Multiscale Image Analysis for the Automated Localisation of Taxonomic Landmark Points and the Identification of Species of Parasitic Wasp

Paul Nathan Angel
School of Computing, The University of Glamorgan

A submission presented in partial fulfilment of the requirements of the University of Glamorgan / Prifysgol Morgannwg for the degree of Doctor of Philosophy.

This research programme has been carried out in collaboration with the Cardiff School of Biosciences and Insect Investigations Limited

October 1999

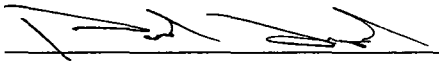
Abstract

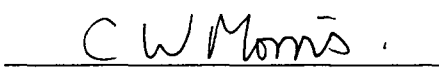
Automating the identification of biological specimens from 2D image data poses difficult problems given the natural variation, specimen damage and background clutter that can exist. The tools used by taxonomists tend to be manual or semi-automated, where the operator locates salient image features from which the system automatically derives taxonomic measurements for identification. Fully automating the extraction of taxonomic features and the subsequent identification task would allow for more robust and accurate identification and provide tools for users in the field who do not possess expert knowledge. This work focuses on the automatic localisation of taxonomic landmark points and the identification of species of parasitic wasps of the order *Hymenoptera* using SEM images of their heads. These images present significant analysis problems. Image feature extraction techniques investigated to solve this problem include deformable contour models, texture analysis and the Mallat wavelet transform. Deformable contour models perform poorly given the textural clutter in the images while texture analysis techniques introduce correlated noise into the segmented image, which can reduce landmark localisation accuracy to 25%. The wavelet transform overcomes this problem by filtering textural clutter at larger scales of analysis. A novel technique is presented which recombines the wavelet transform to create a single contour map where textural clutter is filtered out. This is based on the interaction between edge events which is calculated within a region of interest (ROI) that expands as the scale decreases. In configuring the ROI, a balance must be achieved between filtering textural clutter and eroding salient contours. The landmark localisation accuracy is directly related to this ROI expansion. This represents the main contribution to knowledge. A fast expansion at the high end of the scale range results in a landmark localisation accuracy of 95%. Applying these landmarks to a neural network classifier results in a 91% correct identification rate. This represents a significant improvement over the 65% identification rate obtained by taxonomists and is robust to landmark displacement as a result of contour erosion.

Declaration

The work presented throughout this thesis, except that which has been explicitly referenced, is solely the work of the author, Paul Nathan Angel, and has not been submitted in part or in whole for any other academic award or to any other academic institution.

Signed

 (Author)

 (Director of Studies)

Collaborating Establishments

This research programme was carried out in collaboration with Professor Lynne Boddy of the Cardiff School of Biosciences and Dr. Hassan Dawah of Insect Investigations Limited, Cardiff.

Acknowledgements

I would like to thank my project supervisors Mr. Colin Morris and Professor Bryan Jones along with Professor Lynne Boddy and Dr. Hassan Dawah for their help and guidance throughout this project. I would especially like to thank Hassan for providing the wasp head photographs used in this work. Special thanks also go to my family and to Sian for their patience and support during the last three years.

Contents

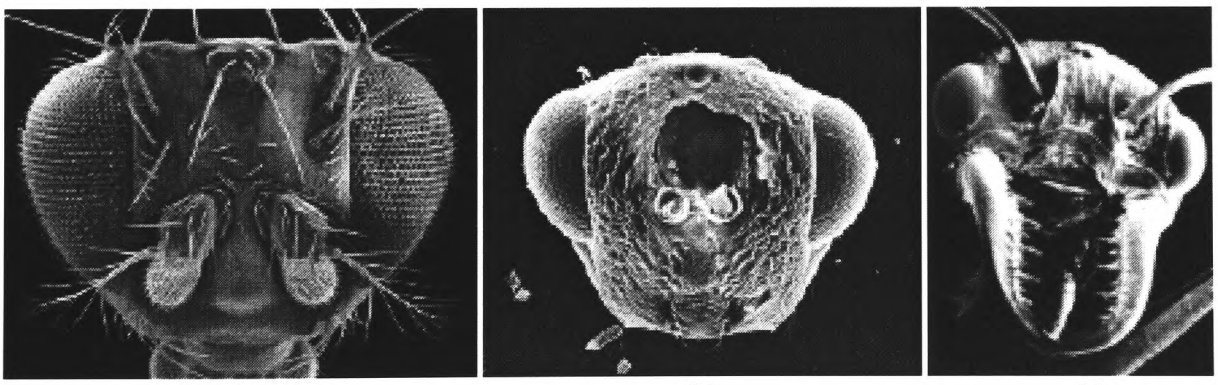
Chapter 1

Introduction

1.1. Introduction

The problems of automating biological classification have become prevalent in recent years. This is due to the increased interest in the application of computational techniques to solve taxonomic problems and the lack of expertise in the respective fields due to the extensive training and knowledge required [11]. The statistical techniques utilised in biological taxonomy [11,35,40] lend themselves towards a computerised implementation. Artificial neural networks, which are also based on statistical theory, are capable of accurate classification and are robust given variable and noisy input data [32]. Once trained, neural networks are also fast in their execution.

Given the diversity of biological taxonomy, suitable representations are required for each problem domain. For example, the analysis of phytoplankton uses a feature vector derived from the spectral signature of specimens processed by a flow cytometer [148]. The extraction of taxonomic landmark points or the analysis of shape structure uses image data as input [15,29,58]. Success in the application of image analysis techniques to medical imaging problems [73,90,96,100,114,120,129,130,143] illustrates their usefulness in the analysis of biological image data, which shares the same problems as medical images : (i) natural variation of the structures in the image being analysed, (ii) specimen damage and (iii) unwanted background clutter [133]. The latter are particularly prevalent for certain types of biological image data where significant textural clutter and specimen damage can exist, as shown in figure 1.1. These problems thus provide sufficient scope for further investigation into the analysis of biological images and will form the basis of this thesis.



(a) (b) (c)

Fig 1.1. Biological images containing background clutter, specimen damage and natural variation. (a) Anterior view of Fruitfly head (*Drosophila melanogaster*). Copyright Dennis Kunkel PhD. University of Hawaii. (b) Anterior view of head of parasitic wasp (*Tetramesa fulvicollis*). (c) SEM image of a Bull Ant. Copyright Tony Romeo, Electron Microscope Unit, University of Sydney.

The aim of this work is to devise a system that will automatically classify parasitic wasps of the order *Hymenoptera* given scanning electron microscope (SEM) images of the anterior view of their heads. To achieve this, it is important to accurately locate the taxonomic landmark points on each head which are then used for species classification. Examples of such images and their associated landmarks are shown in figure 1.2.

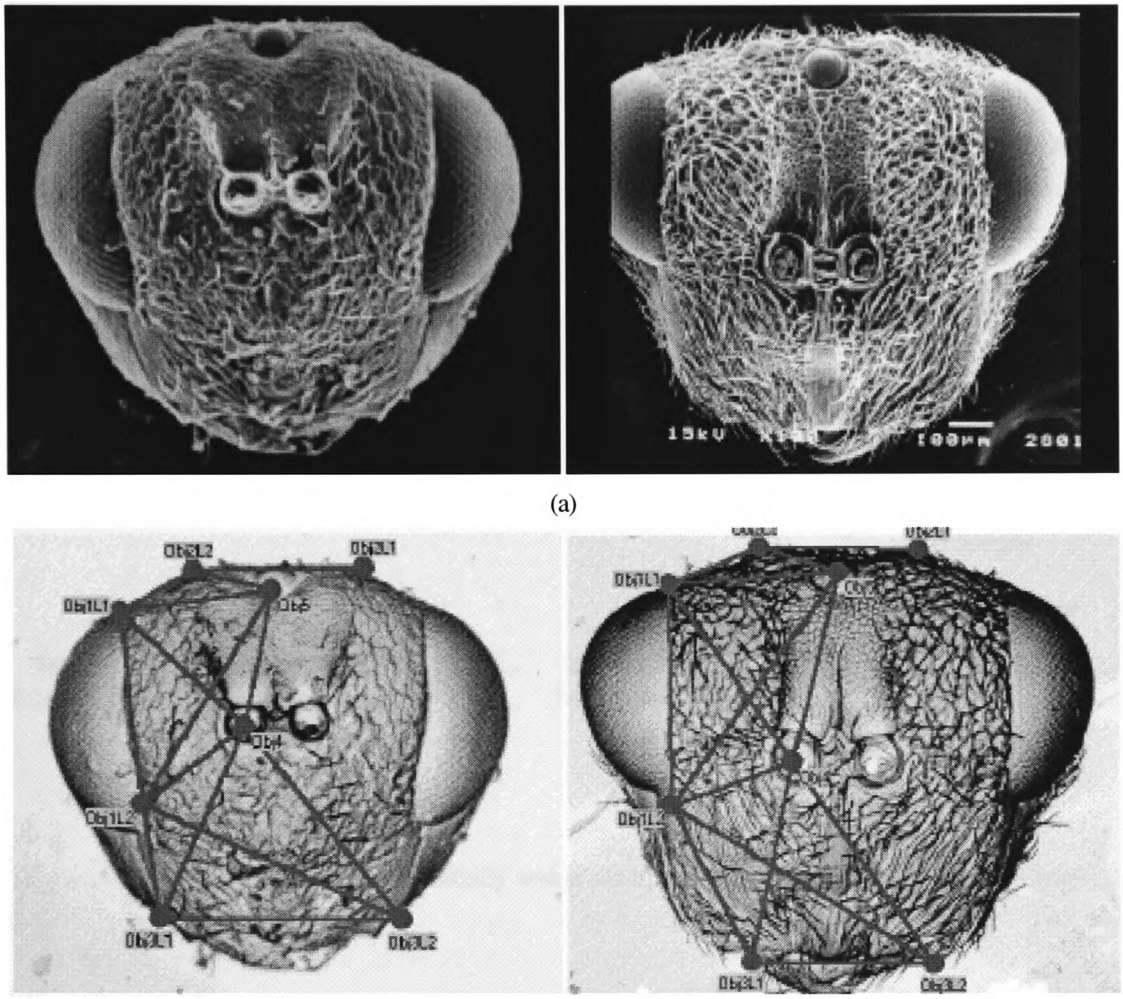


Fig 1.2. (a) Example SEM images of the anterior view of parasitic wasp heads from species Tetramesa linearis and Tetramesa eximia. (b) Landmark features which must be localised for species classification.

Discussions with experts in the taxonomy of parasitic wasps reveals these landmarks to be the salient taxonomic features for species identification. A description of the taxonomic features used to classify each wasp head can be found in appendix 2. Much work has been done on identification using features from wing structures [145]. However, there is not always sufficient information on wing structures to provide an adequate distinction between species,

and so other features are sought, even behavioural and ecological features [21]. From the image analysis perspective, these images present important problems for locating landmark points given the severe deformation and textural clutter that is present. For biologists, there are many reasons why the classification of these species is important. There are currently around 50,000 described species and taxonomists believe this represents only 25% of the total number. Such parasitic wasps are important for ecological stability, but biologists believe many species are becoming extinct before they can be documented. This problem is compounded by the fact that there are very few skilled experts available to perform the traditionally manual tasks of landmark localisation [11], taxonomic feature extraction and species identification given the extensive training required. Aside from academic interests, economically it is important to identify which species a particular specimen belongs to, since some species can cause severe crop damage [20]. A system which incorporates an experts knowledge and can be used by anyone in the field would be most beneficial. For this work, a library of 43 images covering 5 *Tetramesa* species (*T. linearis*, *T. hyalipennis*, *T. calamagrostidis*, *T. fulvicollis*, *T. eximia*) belonging to the *Eurytomidae* family of the order *Hymenoptera* are used. Each image is scanned at 130 dpi. and are 512 x 400 pixels in size. These are shown in appendix 1 while figure 1.3 illustrates the taxonomic hierarchy.

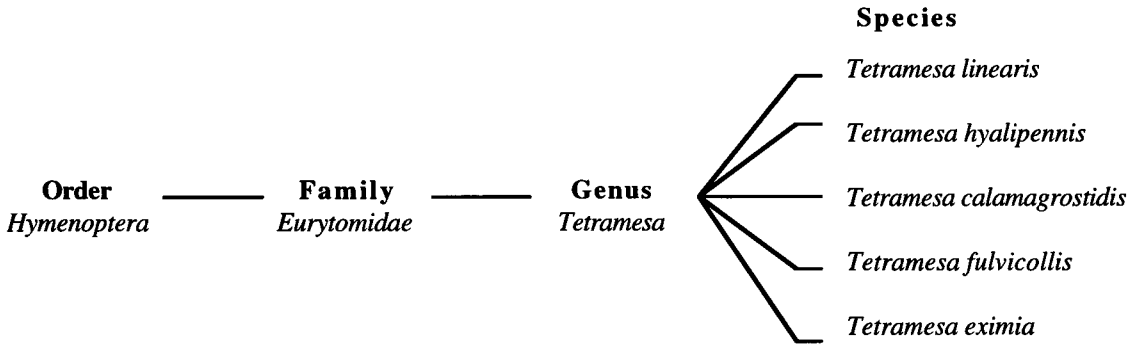


Fig 1.3. Taxonomic hierarchy showing the family and related Tetramesa species which form the basis of this study.

In order to achieve the main goal of this work which is to automatically locate taxonomic landmark points and automatically identify species of parasitic wasp based on these features, the following sub-goals are defined...

• To accurately extract salient image features for representing and localising the taxonomic landmark points.

• To Build suitable models that can encode the deformation present in the wasp head image data.

• Implement robust techniques to localise landmark points

• Implement robust techniques for the automated identification of wasp species.

Automating the extraction of landmark points requires the image to be decomposed and represented in a meaningful way. Mathematically an image is a vector in a high dimensional space and it is the aim of any image analysis system to derive a high level semantic representation from the raw image data [74,122]. Any reasoning done on the image is carried out at this higher level as the high dimensionality of the raw image data is restrictive [122]. As mentioned above, in order to localise landmark points, the issues of irrelevant image clutter and natural deformation, or local shape distortion are of particular interest, and will form the cornerstone of this work. Since the field of image analysis has grown significantly with the emergence of many techniques, this work will only investigate those techniques which are relevant to the problems at hand.

Given the significant textural clutter present in the wasp head images, image feature extraction techniques will focus upon texture analysis and multi-scale edge detection methods [34,78,79,82]. Emphasis will be placed on wavelet techniques since these retain local image properties necessary for landmark localisation. Given the natural deformation within the wasp head images, techniques for modelling and localising deformable shapes will also be studied in depth. Here both deformable contours (Snakes) and curvature scale space are considered for modelling the local object deformations present in the wasp head images. Problems involving the global object transformations rotation and scaling [122] are not addressed. Translation invariance is inherent in the shape representation techniques adopted and scale is considered only implicitly as the local deformation between each wasp head instance. The later discussion on future work will look at how the techniques described here can be augmented to include global rotation and scaling transformations. These approaches represent a "bottom-up" solution, deriving meaningful regions and deformable shapes from the image data respectively. It is widely acknowledged throughout the literature that high level knowledge can be used to guide such low level feature extraction processes [122]. Techniques based on this methodology are referred to as "top-down" approaches. Given the presence of deformation in the wasp head images, hierarchical representations for feature extraction are also studied and applied to the wasp head problem. Other techniques for object localisation that are considered include elastic graph matching and the generalised Hough transform. Face recognition techniques are also studied, as the literature contains numerous techniques for locating salient feature points on human heads. The derivation of the relevant taxonomic features and subsequent species identification are then implemented using a neural network architecture. The presence of irrelevant clutter, or correlated noise in an input image results in structural ambiguity which can give rise to poor landmark accuracy. A novel technique is presented

which filters out textural clutter and retains salient contour features for accurate landmark localisation. This represents the original contribution to knowledge from this work.

The thesis is organised as follows. Chapter 2 discusses the numerous techniques used in image analysis to extract features and represent shape data for both intermediate and high levels of processing. The relevance of these techniques to solving the landmark localisation problem are also considered. Chapters 3 to 7 discuss the implementation of a generic system for extracting features from the wasp head images and locating the taxonomic landmark points. Chapter 4 describes a novel technique for delineating contour and textural features, allowing textural clutter to be filtered out while retaining salient edge features. This is shown to be important for accurate landmark localisation and represents the contribution to knowledge for this work. The results of each stage applied to the wasp head image data are also discussed in detail. Chapter 8 describes the results of using a back propagation neural network to classify each wasp head image using the landmark features which have been extracted. The thesis then concludes with a discussion of the techniques used to solve the landmark localisation problem and how they overcome the limitations of techniques described throughout the literature. Finally, chapter 10 discusses how the techniques implemented can be extended to other problem domains.

Chapter 2

Literature Survey

2.1. Introduction

Traditionally image analysis can be broken down into three stages, low level processing, feature extraction and classification [73,137]. For this work we want to extract salient image features, derive the taxonomic landmark locations and then use these for species classification. Techniques developed over the last few decades have been applied successfully to many image analysis tasks, from simple machine vision systems in controlled environments to the analysis of biomedical images of brain tissue and cell structures [30,133]. Although medical image analysis has received a great deal of attention throughout the literature, entomology presents many new problems due to the complexity of the images. Problems such as background clutter, specimen damage and natural variation can occur within individual species, as shown in appendix 1. Generally, the analysis of biomedical images requires more advanced techniques than the analysis of artificial objects [112]. Hence this research has focused on those techniques which have been applied to problems of biomedical image analysis and are relevant to the landmark localisation problem described above. These include texture analysis, deformable contour modelling and multiscale edge extraction and representation. High level recognition techniques are also considered, including optimisation approaches such as elastic graph matching and statistical based techniques such as the Hough transform. Although conventional bottom-up processes are studied, where images features are extracted and applied to a classifier or high level reasoning process, top down approaches are also considered. Here knowledge of the problem domain is used to facilitate the extraction of image features [129]. The application of hierarchical representations to top down image analysis is studied. Such techniques have been used by Suetens *et al.* [130] for modelling and extracting coronary blood vessels from 2D images. Neural network architectures are also studied for species classification given their robust handling of variable and noisy input data. Face recognition techniques are also studied, as the localisation of feature points is relevant here, too. The feasibility of each technique to solve the landmark localisation and species classification problems are considered. Given the advantages and inherent limitations of each technique described throughout the literature, only those techniques relevant to solving the landmark localisation problem described above are selected for implementation and further study.

2.2. Texture Analysis

Segmentation techniques build abstractions of image regions by grouping pixels that satisfy given homogeneity criteria. Given natural images, colour segmentation can prove useful, but for the images analysed here, colour data provides no additional information over grey scale images. Older techniques such as region splitting and merging through histogram manipulation [73] and the sequential labelling of pixels [46] perform well for regions with little variation in

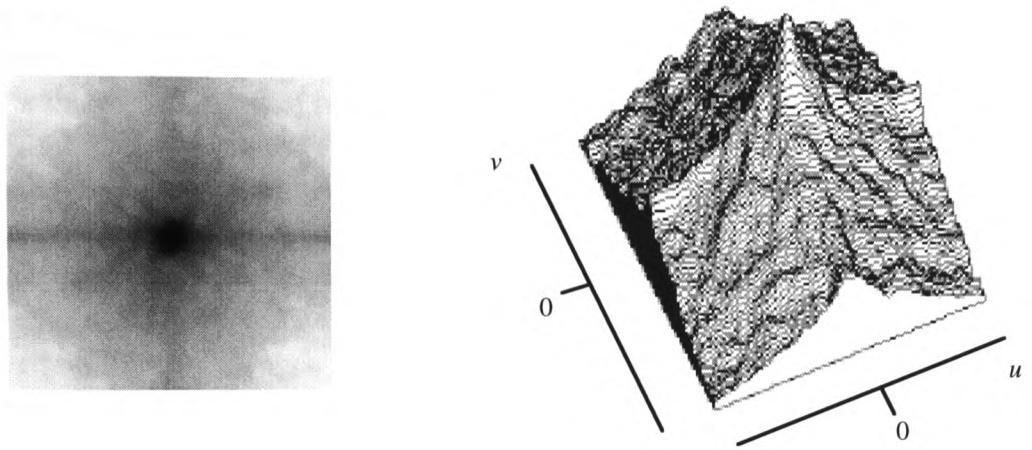
the grey level or statistical properties [143]. Only in cases where the different regions can easily be delineated, as in the case of certain cell images [100], can global histogram approaches be applied. Seeded region growing [1] builds abstractions of regions starting from individual pixels, or seed points which grow subsuming the surrounding region as long as the criteria for homogeneity is satisfied. The problem of placing the seed points must be overcome manually or with high level knowledge about the images being analysed. Dunn *et al.* [29] use region growing to segment TEM (Transmission Electron Micrograph) images of insect hemocytes. Mui *et al.* [90] also use segmentation based techniques to analyse neutrophil cell images. However, this technique works best when regions are continuous. In analysing biological images such as those in appendix 1, global segmentation and region growing techniques will fail. This is because of the significant grey level variation that defines the textural features within each image region. As such, research in this area has focused on texture analysis which encompasses both segmentation and classification, rather than the more traditional approaches to image segmentation described above.

Yu and Berthod have applied game theory to the segmentation of textured images [152]. Their results show that this technique performs well but the segmentation results are sub-optimal. Pyramid algorithms have also been used successfully in segmenting textural images [54]. Although suitable for segmentation, little has been done on using pyramid architectures for classifying texture. Fractal techniques [80] have been used to analyse texture in natural scenes. The fractal dimensionality of a texture can be calculated which reflects only the complexity of the texture. The problem with such an approach is that two distinct textures can have the same fractal dimension because they are equal in complexity. This limitation can be overcome by using a multi-fractal extension which incorporates higher order statistics [30].

Structural, or syntactic techniques have been applied to the analysis of texture [16]. These approaches are based on formal languages where images are constructed out of textural-primitives, or texels through rules that describe the texture [73]. Such approaches work well for deterministic textures, but produce poor results given the natural variability present in biological images. In a related approach, Julesz, a well known psychological and theoretical researcher, considered that the building blocks of texture were dots, elongated blobs and terminations of lines. These are referred to as "texons". Instead of a formal language describing the structure of texture, feature detectors inspired by those observed in the mammalian visual system are used to extract these features and their response used to describe the texture [42]. Given the complexity of the texture in the wasp head images, the resulting representation would be equally complex. Statistical techniques have also been used for texture analysis. One of the more popular techniques is the Spatial Grey Level Dependence Matrix (SGLDM), or co-occurrence matrix [73]. This actually consists of a set of matrices $S_{d,\theta}(i,j)$ derived from the input image showing the number of times grey level i is orientated with respect to grey level j for each d (distance) and θ (orientation). The problem with this approach is that it is

computationally expensive. Another statistical technique based on the work of Chen *et al.* uses a set of Statistical Geometrical Features (SGFs). These are derived from the connectivity of regions obtained from a set of binary images which are extracted from the original textured image. They report that 16 SGFs were extracted which gave a correct classification rate of 86% given a database of 112 Brodatz textures. This shows considerable improvement over earlier statistical techniques [16]. Chen *et al.*'s approach is also scalable, with performance decreasing at a slower rate than other statistical techniques as the data set increased in size. Texture can also be represented as a set of micro edges [25,95] and the statistical distribution of these edges can be used to segment textured images [95]. The heuristic nature of statistical approaches limits their discrimination ability [16] and Unser [140] notes that given they can analyse only a small region within the image they are best suited to micro-textures. Given the natural textural variation within the SEM images being studied, it is important we discriminate between a wide variety of possible features and over a large spatial range. Hence statistical approaches do not lend themselves to the analysis of the wasp head images in appendix 1.

Spectral analysis of textural images has been used extensively in image processing research because the frequency components of the image can be decomposed using a suitable technique such as the Fast Fourier Transform (FFT) [73,112] or the Fast Hartly Transform (FHT) [73]. Each component of the texture can be plotted in the frequency domain, a 2D space showing the frequency and amplitude of each component. This can be viewed as a power spectrum with lower frequencies at the centre and higher frequency components further from the centre. Most of the information in images is contained in low frequency components and so most information is concentrated at the centre of the power spectrum, as shown in figure 2.1. Fourier based techniques are good for periodic texture, but their performance degrades as the periodicity degrades [16]. Such techniques also sacrifice local image properties by analysing the global frequency components of the image. This is because the underlying sinusoidal bases (sine and cosine) do not provide compact support, that is, sin(x) and cos(x) will oscillate across the interval [-1,1] as x tends to infinity. Since the taxonomic landmark points on the wasp head are localised features, it is important that local image features are retained. As such, techniques which analyse the global properties of an image are not relevant.



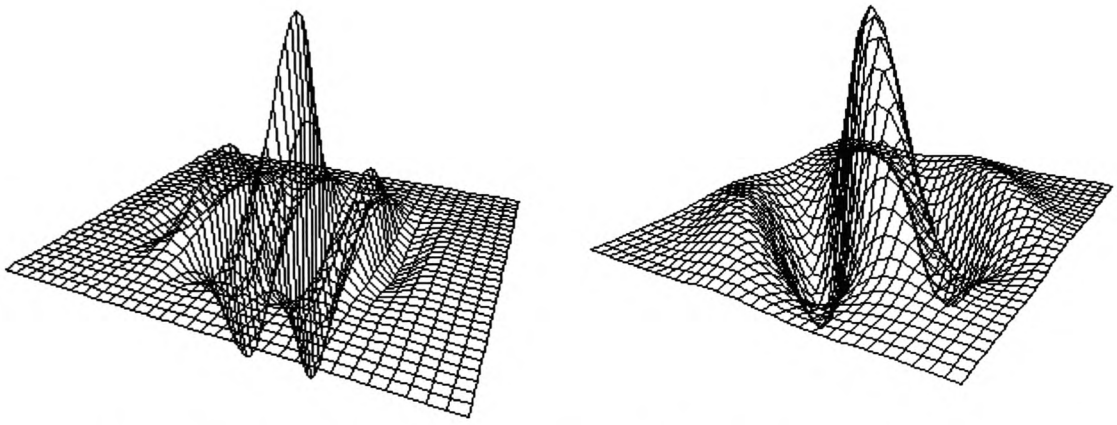
(a) (b)

Fig 2.1. (a) Frequency domain representation of an image. (b) Corresponding 3D surface plot illustrating the concentration of image data around lower frequencies at the centre of the frequency domain.

Due to the power of spectral techniques and the neurophysiological evidence which shows that the human visual system is sensitive to different spatial frequencies [42], much research in texture analysis has focused on conjoint spatial-spectral representations. In such representations, a grid, or lattice of uniformly spaced points is laid over the image and the spectral properties of the image in the locale of each point are extracted. This overcomes the problem of sacrificing local image features by analysing the global frequency components of an image. The Windowed Fourier Transform (WFT) analyses the properties of regularly spaced regions of interest within an image, assigning a strength, or relevance to the points in each region which decays, usually exponentially, from the centre. Wavelet analysis allows the local properties of images to be analysed at the appropriate scale. Unlike the sine and cosine bases of the Fourier transform, wavelets have compact support and thus lend themselves to the local analysis of image data, more so than the WFT approach [19]. For texture analysis, the tree structured wavelet transform, Unser's wavelet frames and the Gabor wavelet, a special case of the Morlet wavelet [156] have been used extensively. The tree structured wavelet transform decomposes different regions of the image based on image content. However, this analyses the global properties of images and is not suited to local texture analysis [97]. Unser's wavelet frames use an over-complete, or redundant representation based on a bank of compact discrete wavelet functions [142]. This representation increases computational expense first because the filter size grows when the effective spatial extent of the filter does not and second because more processing is carried out to produce redundant information [97]. Research into the mammalian visual system has shown that the receptive field of certain cortical cells approximate the Gabor wavelet very closely [23,49]. In considering texture, the Gabor, which is defined as a gaussian modulated sinusoid [118], can be configured to a particular orientation and spatial frequency. Daugman [23] has shown that the Gabor achieves a lower bound on the joint

entropy which is defined as the product of effective spatial extent and frequency bandwidth. This means that Gabor functions provide sufficient spectral information at each point along a signal, or image. Examples of two dimensional Gabor functions are shown in figure 2.2.

Daugman [23] and Porat and Zeevi [98,99] have used the Gabor to model images. An image can be represented by a set of 2D Gabor functions positioned on a two dimensional lattice, each configured with a set of coefficients. To find these coefficients Daugman uses a neural network architecture while Porat and Zeevi consider Gram's minimum distance calculation. From the calculated coefficients, Daugman uses the local spectral signature of the texture to delineate and classify each region while Porat and Zeevi derive six localised features based on frequency, orientation and intensity information [99]. Both approaches report good results. Also, Earnshaw shows that the Gabor wavelet is suitable for analysing texture in both artificial and natural images [31].



(a) Gabor: Frequency=1, orientation=0^{o} (b) Gabor: Frequency=0.5, orientation=45^{o}

Fig 2.2. Two dimensional Gabor wavelets.

Generally, extracting features from an image using the Gabor wavelet involves convolving a bank of filters configured at different orientations and frequencies. Daugman shows that the frequency range of the filter bank should adequately cover the frequency domain [23]. Texture can then be analysed based on the response of each filter. The problem with this approach is that it is computationally expensive for a large number of filters and the performance degrades if the frequency and orientation of the texture does not correlate with that of the filter bank. To overcome this Ravichandran and Trivedi have developed a filter based on circular-mellin features [106], an example of which is shown in figure 2.3. Unlike the Gabor, it is less sensitive to changes in the orientation of the texture. Although this is suitable for segmentation, a bank of filters like that used for the Gabor wavelet is still required.

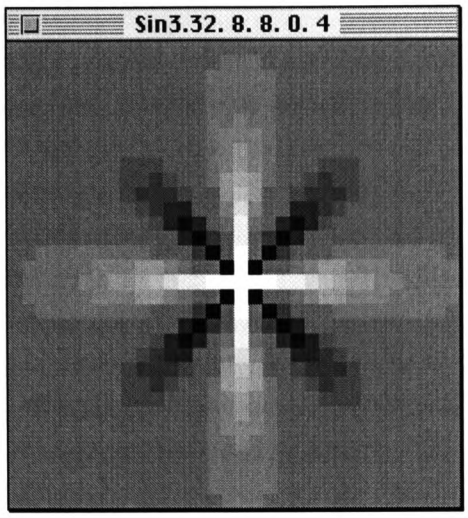


Fig 2.3. Circular-Mellin function showing characteristic spokes.

Reed and Wechsler [107,108] propose a similar conjoint scheme using the Wigner-Ville (W-V) transform. This is improved upon by Zhu *et al.* [157] who calculate a feature image from the smoothed W-V transform. Since the local spectral plane is symmetrical about its origin, this approach suffers from interference between positive and negative local spatial frequencies. Also, local image properties can cause interference in the spectral domain. This affects the resulting feature image, impairing segmentation capability. Zhu *et al.* [157] propose methods to overcome these limitations, improving results over Gabor-based approaches for artificial images [156,157]. The segmentation capability is not improved for natural images however.

Optimal approaches to texture analysis reduce the number of filters required by extracting the salient textural features and only using those wavelets which are necessary [102]. Various criteria are used to discriminate between texture, including the Unser [140], Fisher and Mahalanobis-Singh criteria [103]. Randen and Husøy show that these criteria are optimal for texture discrimination and that combining them produces an improvement over the use of each criterion in isolation [103]. They show that this unified approach produces significantly better results and that these criteria are superior for texture discrimination [103]. Shang and Brown [117] utilise a neural network architecture to reduce the dimensionality of the texture image by extracting salient features which are then used by a classifier network. Their results are good, but a large number of training images are required, typically 5000 training patterns for good segmentation. Lu and Xu [76] also implement a neural network to classify textures, with good results on natural textures. However, we are likely to encounter unseen textural patterns which can vary significantly from one species to the next. For example, texture varies significantly between species *Tetramesa fulvicollis* and *Tetramesa eximia*, instances of the former being used for the training of shape data, described later, while the latter is unseen during training. Indeed, Lu and Xu state that segmentation will not work for all textured images given the

difficulties in obtaining a good texture representation [76]. This is particularly the case for natural textures which can vary significantly. The Adaptive Gabor filter has also been developed where the frequency and orientation of each filter are configured according to dominant frequency and orientation of the texture [97]. Using such optimised texture analysis reduces the computational complexity of the system, but does not take into account the textural variance that is present in natural images, as shown in appendix 1. Texture analysis techniques suffer from the problems of poor edge localisation and the introduction of erroneous regions [102]. This is due to natural variation in the texture and is a problem that is relevant here. However, despite the fact that accurate edge localisation is vital for locating the taxonomic landmarks on each wasp head, the textural content of the SEM images being studied warrants further investigation of texture segmentation techniques. Table 2.1 summarises the texture analysis techniques discussed above.

Table 2.1. Summary of Texture Analysis Techniques

Technique	Useful	Reason
Game Theory	No	Sub-optimal segmentation
Fractal Complexity	No	Two distinct textures can have the same fractal complexity
Structural / Syntactic	No	Only useful for deterministic textures - Wasp head texture changes as a result of natural deformation
Statistical	No	Poor discrimination ability and can only analyse a small region in the image
Fourier Analysis	No	Sacrifices local properties which are necessary for localising landmark features
Wavelets (Gabor)	Yes	Retains local properties and provides good natural texture segmentation
Optimised Wavelet Approaches	No	Require textures to be known *a-priori*. Cannot assume all natural texture variations on the wasp head are known.

2.3. Edge / Contour Detection

Edge detection is the converse of region segmentation in the sense that we are looking for points in the image that delineate regions rather than points that combine to form complete regions. Edges, or contours contain invaluable information about the shape of objects in an image. Attneave observed that most of the information about object shape is contained at edge points of high curvature [6]. As a result, a great deal of research has been applied to extracting edge information and using this to describe and classify objects in a given image. Edges can be considered sharp changes or discontinuities in the grey level of an image. An idealised edge is referred to as a "step-edge", shown in figure 2.4(a), but the configuration of edges is not limited to this case with ramp edges, shown in figure 2.4(b) and noisy edges, shown in figure 2.4(c) delineating regions within an image. The latter is more common due to environmental conditions and hardware factors that introduce noise into images.

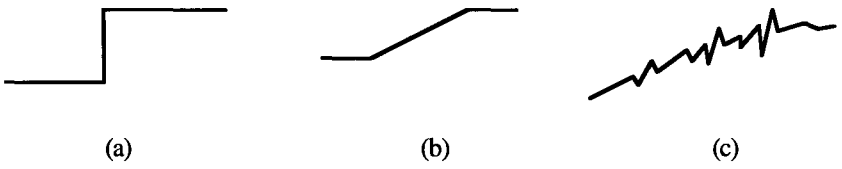


Fig 2.4. Different edge configurations.

Mathematically, images can be considered two dimensional functions, or signals. Edge detection research has focused on the development of differential operators, or filters that are convolved with the image at each point to enhance sharp changes in the gradient of the image [73]. Canny considers edge detection as an optimisation problem and proposes three criteria by which the performance of an edge detection operator should be judged. These are the *detection* criterion which states that only important edge events should be detected and non-salient events ignored, the *localisation* criterion which states that the location of an edge response should be close to the true location of the corresponding edge and the *one response* criterion which states that a minimal number of responses should be produced for an edge [13].

Generally filters can be classed as high pass or low pass. The former enhances high frequency features while the latter highlights only low frequency components. Early techniques were basic 2D extensions of the simple 1D difference function [112]. One of the earliest approaches was the Roberts Cross operator [109] which was later extended by Sobel [121] to overcome the sensitivity of the operator to edge direction. Kirsch [58] further improved upon Sobel's operator by considering not just the vertical and horizontal derivatives of the image but the derivative of all 8 possible directions at each point (given a square tessellated digitised image). Both the Sobel and Kirsch operators produce good results and are used widely in image processing applications [73].

The Roberts Cross, Sobel and Kirsch filters are first derivative, fixed size operators, producing a maximum at sharp gradient changes in the grey level of an image. Neurophysiological evidence shows that different cells in the retina have different receptive fields, that is different regions respond in different ways when stimulated. Marr [81] has shown that these cells perform edge detection operations.

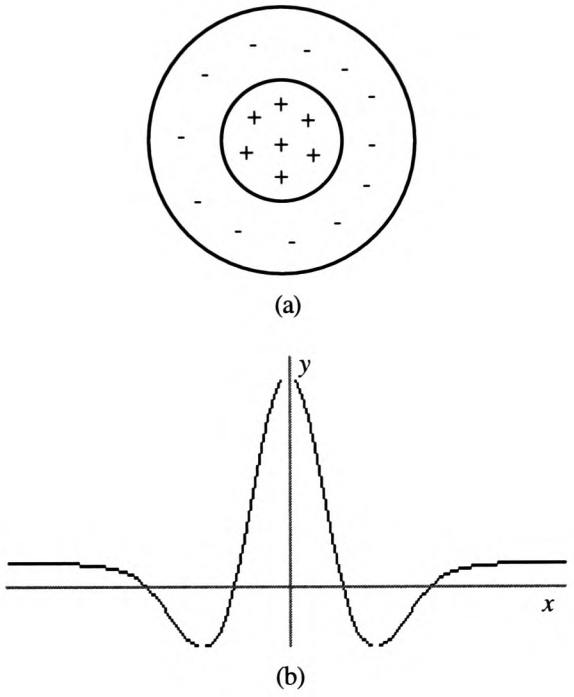


Fig 2.5. The Marr/Hildreth operator.

Figure 2.5(a) shows the receptive field (polarity) of a retinal cell, while figure 2.5(b) shows the profile of this receptive field. This profile is important because it approximates the second derivative of the gaussian smoothing function. This function has become known as the Marr/Hildreth operator [82]. Its more common name is the "Mexican hat", or "Sombrero" filter. Related to this is the Laplacian of Gaussian (LoG) operator which is in fact a special case of the Marr / Hildreth operator, where the width of the central positive region equals one pixel. If the width is greater than one pixel, the filter reduces to the former. Unlike the early filters described above, the Marr/Hildreth operator is a second derivative operator which produces zero crossings in the filter response instead of a maxima. Unlike the fixed size filters, including the LoG operator, the Marr / Hildreth operator is scalable, allowing it to be configured as a high pass or low pass filter. More meaningful information can be extracted from noise corrupted images using scalable operators since noise can be easily filtered out by increasing the width, or scale of the filter function. Fixed size filters, however, are sensitive to noise.

2.4. Multiscale Analysis

Interest in the analysis of image data across different spatial scales has grown over the last decade since more meaning can be extracted from the multiscale representation of an image than can be by looking directly at the raw pixel values [150]. Marsic notes the parallels with human vision, where only relevant features are focused upon according to the current problem [83]. Not all contours in an image can be extracted with a single scale of analysis, and as a result, numerous techniques have been developed to address this problem including pyramid algorithms [73,54], Witkin's scale space [150] and wavelet analysis [23,79,82,86,98], the latter of which has only recently brought together many mathematical notions of multi-scale analysis into a coherent theory [22].

2.4.1. Pyramid Algorithms

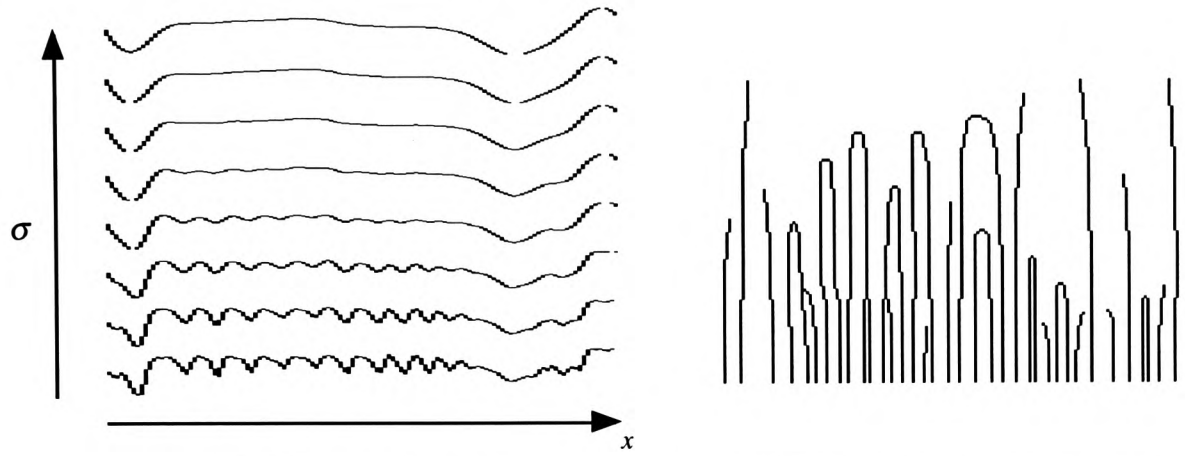
Pyramid algorithms were developed as a way of deriving a multiscale representation of an image. They work by creating a series of progressively smaller images, the pixels in each image representing a region, or group of pixels in the larger image upon which it is based. This process continues until a 1x1 pixel image remains. The construction of such pyramids depends on their decimation ratio, or the rate at which the size of the image decays from one level to the next. Typically the decimation rate is constant, but newer architectures referred to as Adaptive Pyramids have a variable decimation ratio based on the content of the image being analysed. Pyramid algorithms have been used successfully in extracting contours and delineating regions of varying complexity [54].

2.4.2. Scale Space Analysis

Witkin's scale space comprises a set of smoothed continuous one dimensional signals produced by convolving an input signal with a smoothing function of increasing width [150]. For an input signal f(x) and a scaling parameter σ, the scale space decomposition $L(x,\sigma)$ is defined as

$$L(x,\sigma) = \int_{-\infty}^{\infty} f(u) \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-u)^2}{2\sigma^2}} du \qquad (1)$$

Figure 2.6(a) illustrates the decomposition of an input signal across different scales.



(a) Scale space images L(x,σ). (b) Event contours across scale space

Fig 2.6. Scale space representation of an image.

Edge "events" [150] in the image can be tracked across scale space by analysing the progression of differential singularities [33], maxima or zero crossings produced by convolving the input signal with the first or second derivative of the smoothing function respectively. These events correspond to inflection points in the smoothed images. As the scale increases, the position of events changes due to the blurring of the smoothing function and its derivatives. Eventually, two events will converge at the apex of characteristic scale space arches where they will annihilate each other. Linking corresponding events across the scale range highlights the event contours that define these scale space arches as illustrated in figure 2.6(b). The symmetrical and convolution properties of the gaussian function make it suitable for scale space decomposition [150], and the resulting scale space retains two important properties. These are the Causality property and the semigroup property. The causality property states that as the scale space is traversed from coarse to fine scales, that is σ decreases, differential singularities can emerge but cannot disappear. Hence all events can be tracked to the smallest scale. The semigroup property [70,91] states that two consecutive smoothing steps with a given gaussian kernel is equivalent to one smoothing step with a larger kernel. As a result, each level in scale space is related to previous levels. Witkin characterises events by their position in scale space, at the apex of the respective scale space arch, and their true location, which can be found by following the event contour to the smallest scale of analysis. The more important and stable the event, the further along the scale axis it resides [150].

Discretisation of Witkin's continuous theory for one dimensional signals requires careful consideration regarding the convolution kernel. A common technique is to discretise the

convolution integral given in equation 1, resulting in a discrete approximation defined as

$$L(x,\sigma) \approx \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-n^2}{2\sigma}} f_{in}(x-n) \qquad (2)$$

However, this can violate the causality and semigroup properties as demonstrated by Lindeberg [70]. Formally, the semigroup requirement specifies that a scale space response $L(\bullet,\sigma_2)$ can be derived from two consecutive smoothing steps $L(\bullet,\sigma_0+\sigma_1)$ where $\sigma_0,\sigma_1 < \sigma_2$. Using the discrete form of the convolution integral, a valid scale space cannot be produced when σ_0 or σ_1 is greater than zero. Hence, a valid scale space can only be produced when $\sigma_0=\sigma_1=0$, so $L(\bullet,\sigma_0+\sigma_1)=L(\bullet,0)=f$, or the ratio $\sigma_2/\sigma_0+\sigma_1$ is an odd integer [70]. For the discrete implementation of the continuous theory it is recommended that the discrete diffusion equation defined in (3) be applied [70].

$$L(x,\sigma) = \sum_{n=-\infty}^{\infty} T(n,\sigma)\, f(x-n) \qquad (3)$$

where

$$T(n,\sigma) = e^{-\alpha\sigma} I_n(\alpha\sigma) \qquad (4)$$

$I_n(\alpha\sigma)$is a modified Bessel function, σ is the scale parameter which is greater than zero and α is a positive scaling factor. Using this method, a discrete analogue of the gaussian function can be obtained and a valid scale space can be derived which retains the causality and semigroup properties of the continuous theory.

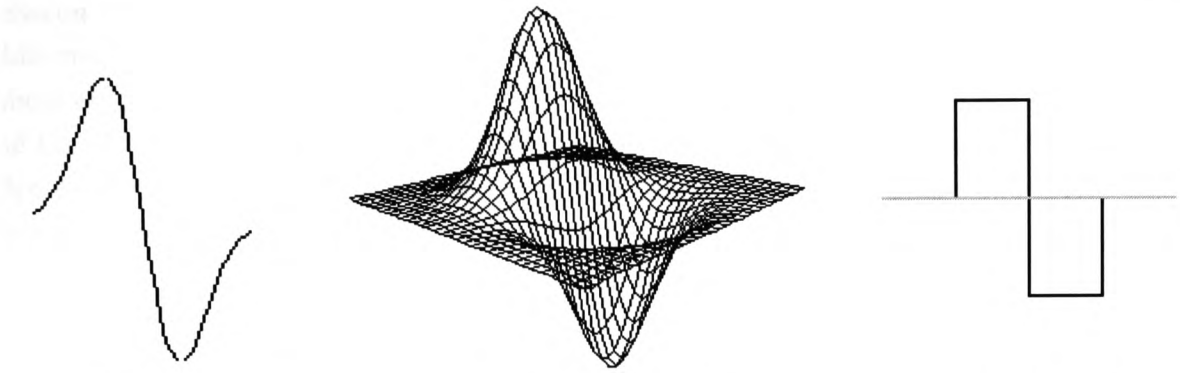
Extending the one dimensional theory to two dimensions introduces new problems. Not only does the location of edge events change across the scale range, but they also split and merge at bifurcation points [33]. In such cases, there is no guarantee that the causality property applies and that any filter has the properties to overcome this [33]. Tracking these events across scale space remains an open research problem. The techniques required are similar to those of surface reconstruction, where points on contour slices are linked to form three dimensional objects. Ulunipar and Medioni [139] consider the problem for one dimensional signals, but do not address the problem of interacting events which occur in active regions of a signal. Thirion and Gourdon [136] have developed a "marching lines" algorithm to track events across contour slices in two dimensional scale space. Mallat and Zhong [79] have implemented an algorithm

for one dimensional signals which links corresponding maxima across adjacent scales based on the sign and proximity of neighbouring edge events. Niessen *et al.* [91] link events across scales within a hyperstack architecture using grey level intensity values. The problem of tracking edge events across scale space is alleviated when an input image is analysed across a dense range of scales. However such an approach is computationally expensive [79]. Hence the choice of scale range is very important. If a fixed set of scales are chosen then the scale space decomposition is equivalent to a pyramid algorithm of fixed decimation ratio. If a continuous range is chosen then arbitrarily dense scales can be selected and configured according to the content of the input image. It is well known in information theory that logarithmic sampling leads to informationally uniform sampling density [33]. Hence a dyadic sequence can be used to retain salient information about a given image [79]. This also reduces the computational requirements and eliminates redundancy since analysis over a dense scale range produces similar results at adjacent scales [79]. However, the problem of tracking events at larger scales remains because of the sparse sampling density.

2.4.3. Wavelet Analysis

Wavelets are mathematical tools that allow functions to be decomposed and analysed locally across a number of scales. A wavelet can be any function $\psi(x)$ that satisfies the condition $\int_{x=-\infty..\infty} \psi(x)\, dx=0$ [78,141]. Wavelet decomposition involves convolving a given image $f(x)$ with dilated and translated versions of a mother, or prototype wavelet, $\psi(x)$. Dilation of the mother wavelet facilitates the multiscale analysis of the input image. Multi-wavelets work in a similar way, only they have multiple scaling functions. These allow important characteristics such as short support, orthogonality, symmetry and vanishing moments to be maintained, although they have been applied mainly to image compression and reconstruction [128]. Different wavelets are suited to different tasks and the Marr/Hildreth operator, commonly referred to as the Marr wavelet [82], the Gabor wavelet [86] and the Mallat wavelet [79] have been widely used in multi-scale edge detection research.

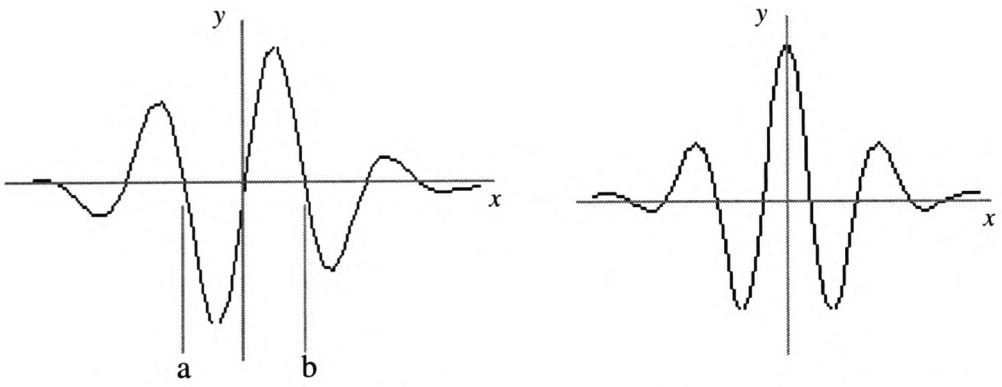
The Marr wavelet described above approximates the second derivative of the gaussian while the Mallat wavelet shown in figures 2.7(a) and 2.7(b) is defined as a quadratic spline function which approximates the first derivative of the gaussian [79]. As noted in the above discussion on Witkin's scale space, the gaussian function has desirable properties for multi-scale analysis. The Haar wavelet can also be used for edge detection, but unlike the Marr and Mallat wavelets, the Haar operator is a square waveform which is discontinuous at x=0. This is shown in figure 2.7(c).



(a) 1D Mallat wavelet (b) 2D Mallat wavelet (c) 1D Haar wavelet

Fig 2.7. The Mallat and Haar wavelets used in edge detection.

The Gabor wavelet can be configured as asymmetrical (sine modulation) or symmetrical (cosine modulation) with 90 degrees phase difference between them. The former can be used for edge detection when bounded across the interval [a,b] as shown in figure 2.8(a) since this region of the function approximates the first derivative of the gaussian [86].



(a) Asymmetrical Gabor (odd) (b) Symmetrical Gabor (even)

Fig 2.8. Asymmetrical and symmetrical Gabor wavelet configurations.

Where the Mallat wavelet produces optimal results based on Canny's criteria [13], the Gabor is sub-optimal [86]. Optimal second derivative operators have also been developed [116], but it is the properties of the first derivative filter responses, local maxima at sharp gradient changes, that have advantages over the zero crossings of the second derivative response. Inflection points in the gaussian smoothed image correspond to the maxima and minima of the first derivative response. Maxima correspond to sharp gradient variations while minima correspond to slow gradient variations. Since zero crossings correspond to both maxima and minima, as shown in figure 2.9, it is difficult to distinguish between these features. In addition, zero crossings cannot differentiate small amplitude fluctuations and as such cannot distinguish

between salient and non-salient features [79]. Throughout the literature it is often the case that additional post-processing is required to eliminate false zero crossings that correspond to non-salient events in the input signal. However, it has been observed by Yuille and Poggio [153] and Lu and Jain [75] that the first derivative gaussian response produces spurious and missing edge points at larger scales. This is in fact a result of the bifurcation points described above, and can result in any gaussian-based filter response.

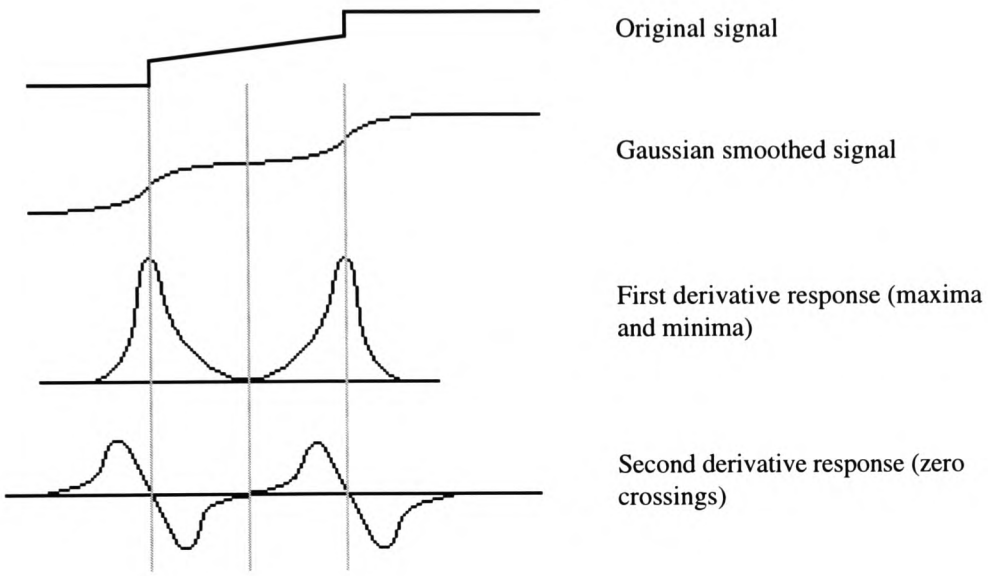


Fig 2.9. Correlation between inflection points, maxima and minima and zero crossings.

Since the Marr wavelet is rotationally symmetric only one convolution is required at each point in the image, whereas the directional wavelets such as the asymmetrical Gabor and Mallat wavelets require two orthogonal convolutions, both horizontally and vertically. This increases the computational requirements but directional filters are better at handling noise present in the input image [112].

Scalable edge operators such as the Marr and Mallat wavelets can be reduced to special cases within Gokmen and Jain's $\lambda\tau$-space, where λ determines the scale, or width of the underlying gaussian curve ($\lambda=\sigma$) and τ determines the continuity around the point $x=0$. When $\tau=0$ an operator is discontinuous at $x=0$ and provides good localisation of edge events but is sensitive to noise. As τ tends to 1, the operator becomes continuous around $x=0$ but localisation of events in the filter response is sacrificed for more robust handling of noise. Gokmen and Jain recommend that filters are configured with a τ value equal to 0.5, representing a balance between edge event displacement and robust handling of noise [41]. Figure 2.10 illustrates the different filter configurations that are possible.

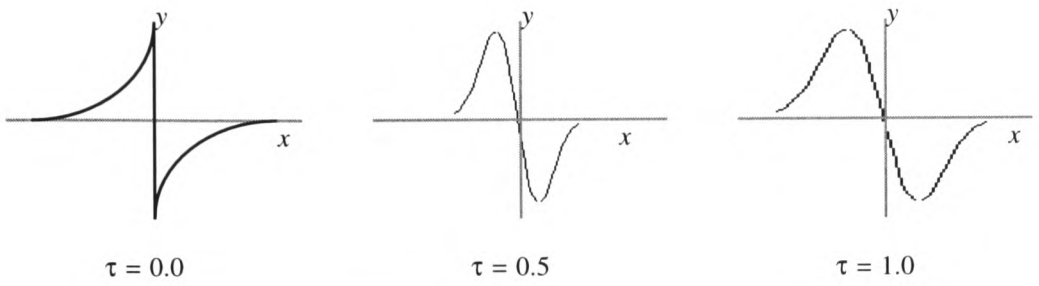


Fig 2.10. Edge detector configurations that can be created in λτ-space.

Wavelet bases are another technique used to extract the necessary information for a multi-resolution description of an image [19]. However, their application to visual pattern recognition is limited because they are not invariant to translation [77]. The information at each resultant scale is also uncorrelated, making the process of linking corresponding edge events difficult [91]. As noted above, multiscale edge representations which are derived from gaussian based operators do not introduce erroneous features at larger scales and the inherent displacement of edge events can be overcome by finding their corresponding events at the smallest scale of analysis. This overcomes the problems of poor edge localisation and the introduction of erroneous regions inherent in texture segmentation approaches. Given this and the fact that taxonomic landmark points lie on the boundaries of structures on the wasp head, multiscale edge detection techniques will be investigated further.

2.4.4. Nonlinear Multiscale Representations

The problem of using a single sized smoothing function for each scale level is that high frequency details in certain areas of the image can be blurred out when ideally they should survive at larger scales of analysis. A great deal of research has gone into developing scale space representations that satisfy the above causality and semigroup properties but also provide additional constraints on blurring, with the size of the kernel being dependent upon the local image data. Jolion & Montanvert's adaptive pyramid [54] which is described above can be considered a nonlinear multiscale representation because the decimation of the image size is dependent upon the actual image data. Two nonlinear multiscale representations have been described throughout the literature [91]. These are the luminance conserving and geometric flow scale spaces. The former retains the grey level values from the original image. The gradient-dependent diffusion equation has been applied by Niessen *et al.* to selectively smooth-out irrelevant structures while edges are retained [91]. Geometric flow scale spaces allow the study of the evolution of curves or surfaces across multiple scales. The Euclidian shortening flow technique has also been implemented by Niessen *et al.* and compared to both the linear multi-scale techniques described above and the gradient-dependent diffusion approach. Their results show that the difference between each approach is not significant, although the

nonlinear approaches require less post-processing [91].

2.4.5. Classifying Edges from Multiscale Information

An important area of research that has recently emerged is the classification of edges based on the progression of events across scale space. Lu and Jain incorporate high level knowledge about the behaviour of different types of edges in scale space [75]. The maxima values contained in the Mallat wavelet transform can be used to derive Lipshitz exponents which characterise different types of edges in the original image [78]. The displacement pattern and stability of edge events in scale space is considered important, with stable events of small spatial extent corresponding to image contours of high curvature. Edge points of high curvature, corners for example, contain significant information about local curve shape and such features are being applied to the problem of automatically identifying landmark points [11,33]. Figure 2.11 illustrates the behaviour of a landmark point located on the top of the wasp head within scale space.

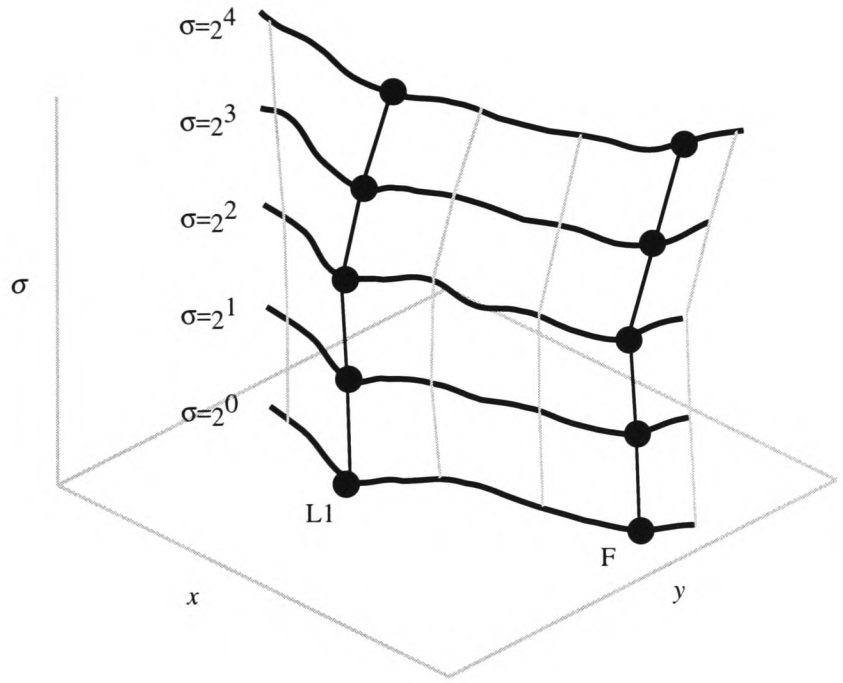


Fig 2.11. The contour segment around landmark point L1 (which in this case is located at the top of the wasp head structure) forms a 3D surface in scale space. The behaviour of this landmark point over scale space is not easily distinguishable from the behaviour of false landmark points (referred to as point F in this example).

Clearly the point does not diverge significantly from its true location, nor does it split and merge with surrounding edge points. The extent of the landmark point in scale space is also significant, surviving at the largest scale of analysis used in this experiment. However,

surrounding edge points can also exhibit these same properties, as illustrated in figure 2.11. Clearly, local edge point behaviour within scale space may be useful for locating local contour features, but significant numbers of false candidate landmarks can emerge in the locale of the true landmark location. So although the behaviour of edge events is useful of locating certain types of contour features within the image, it is not suitable for locating taxonomic landmark points on the wasp head images used in this work. Table 2.2 summarises the edge detection techniques discussed above.

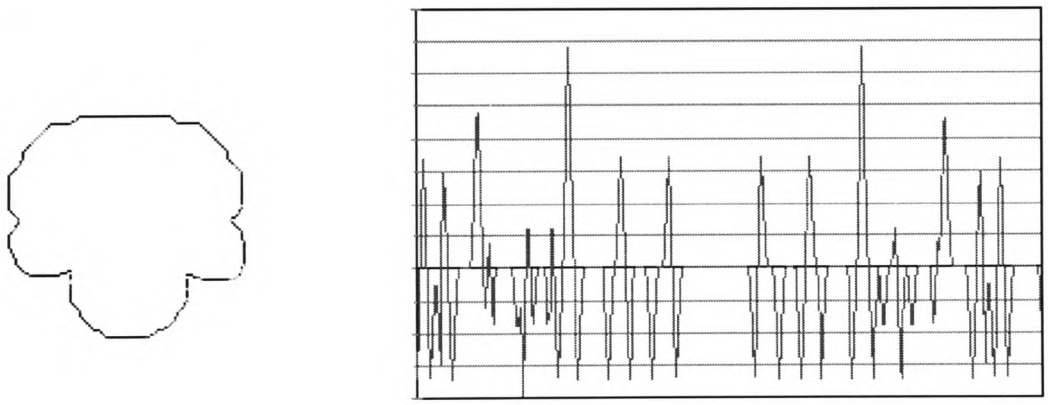
Table 2.2. Summary of Edge detection Techniques

Technique	Useful	Reason
Fixed sized edge detector	No	Sensitive to high frequency image features such as texture and noise
Multiscale Analysis (Wavelets)	Yes	Allows noise and non-salient textural clutter to be filtered out
First Derivative Edge Detectors-Mallat Wavelet	Yes	Allows important edge features to be distinguished from unimportant ones
Second Derivative Edge Detectors-Marr Wavelet	No	Cannot distinguish between important and unimportant edge features
Locating landmarks using multiscale edge behaviour	No	False landmark points can exhibit the same behaviour as true landmark points

2.5. Shape Representation

Numerous representations exist that allow us to describe an object's shape from the image features extracted during the low level feature extraction phase. Since the images in appendix 1 contain significant textural detail, region-based descriptions should be based on textural properties, as described in section 2.2. Since the landmarks of interest are defined as either entire structures or points on structure boundaries, as shown in appendix 3, this section studies those techniques used to model object boundaries, or contours. Freeman's chain coding is a simple method of representation where a contour is defined as a sequence of directional codes describing the local structure of the contour [36]. Given a square tessellated image, each point on the contour can proceed in one of eight possible directions. This is adequate for simple shapes, but higher order chain codes are required for greater flexibility [122]. Syntactic approaches fragment the contour into its component parts by assigning a meaningful label to

each part. This is suitable for rigid forms, where components can be assigned labels such as "corner" and "line", but forms which exhibit natural curvature do not lend themselves to this representation [5]. Eigenshape analysis models shape data using a set of eigenvectors. These represent a series of angular variances between points along the boundary of a given shape which are decomposed to form a matrix of principal axes, or eigenshapes. This is considered by Lohmann to be a generic tool for subsequent morphometric analysis [72]. However, Full and Ehrlich show that the technique is in fact not generic, in some cases being invalid for extracting maximal information from morphometric data. They also show eigenshape analysis to be sub-optimal, leading to misunderstandings of morphological variation [37]. Akra *et al.* [3] consider the problem of extracting salient feature points in an image for matching against stored models. The emphasis of their work is placed on filtering non-salient edge points. Semantic information about feature points, such as taxonomic landmarks, is not considered. Given the curvature properties of a shape's boundary, the points of high curvature, which are important features for locating landmark points as noted above, can be extracted. A planar curve can be represented parametrically as $C(s)=(x(s),y(s))$ where the curvature at each point P on the boundary curve is defined as the rate of change of the angle subtended by the tangent at point P with the x axis. This is equivalent to calculating the curvature with the derivatives of the horizontal and vertical boundary functions [104]. Figure 2.12 shows the corresponding curve function for a 2D contour.

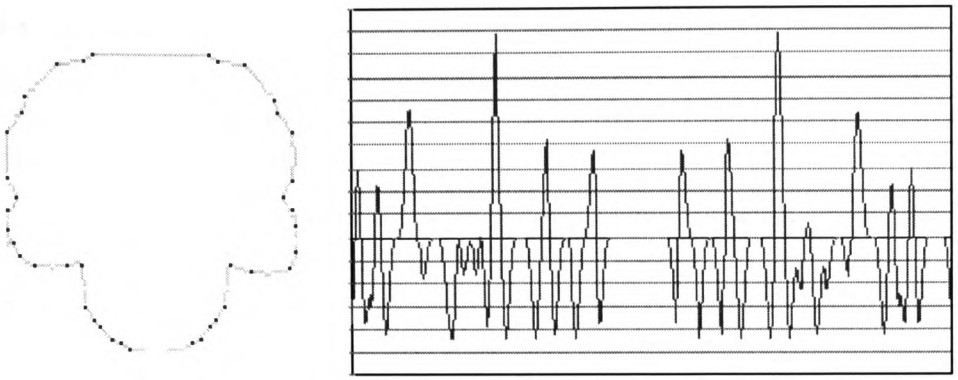


(a) (b)

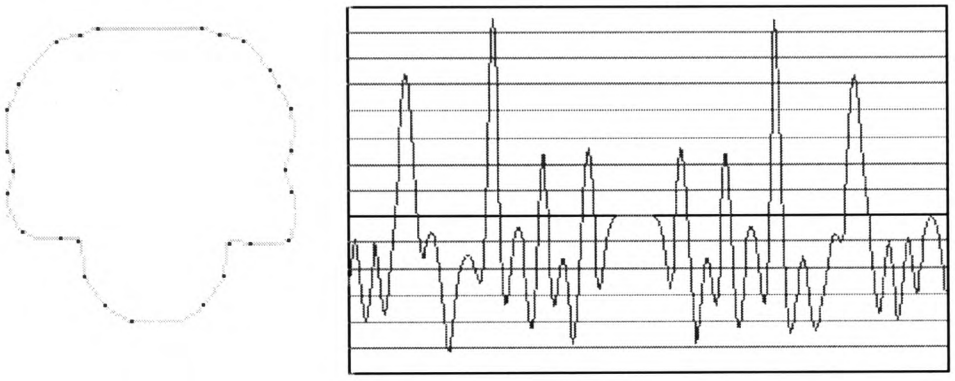
Fig 2.12. (a) 2D planar curve. (b) Resulting curvature function. Peaks correspond to points of high curvature and the sign represents the concavity/convexity of the curve points. From Rattarangsi, A. and Chin, R.T. "Scale-Based Detection of Corners of Planar Curves". IEEE PAMI 1992.

Numerous algorithms exist to find dominant curvature points, or "corner points", but they only work well for objects which have similar sized curve features and information about the object size is known [104]. The problem of extracting dominant curvature points from planar curves is that boundary noise can introduce false corner points. Curvature scale space, based on the

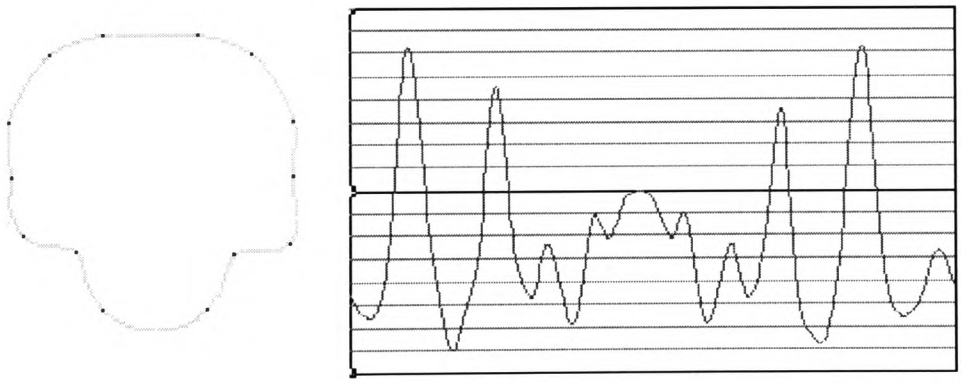
same principles as Witkin's theory described above, allows the global properties of planar curves to be represented and overcomes the problem of boundary noise [104]. Sankar and Sharma propose a variant on multiscale analysis by averaging the sign of the local curvature where +1 represents convex curvature, -1 represents concave curvature and 0 represents no curvature. After a number of iterations, points with maximum average values are considered to be dominant curve points [115]. Teh and Chin propose an algorithm where the region of support for calculating curvature is based on local curve properties. However, this approach is not robust to boundary noise [134]. Saint-Marc and Medioni smooth the curve function while retaining discontinuities which represent corner points [113]. However, an appropriate smoothing parameter is required otherwise corner points are filtered out if the smoothing parameter is too large or false corner points are retained if the smoothing parameter is too small. Dudek and Tsotsos propose a curvature tuned smoothing (CTS) approach where the contour is fragmented into its component parts according to local shape properties and each segment is represented in scale space [27,28]. Object recognition is done by matching segment sequences. However, this does not take into account the part-whole problem, where isolated contour segments in the image plane must be matched with their corresponding sections in a complete curve. Suitable segment ordering would have to be carried out, but this is not considered in their work. Since a multiscale representation generates multiple versions of a function, each with a different degree of detail, Meer *et al.* select the most appropriate scale at which a given shape can be represented [85]. However, salient information does not always reside at a single scale of analysis [81]. This is acknowledged by Rattarangsi and Chin [104] who utilise the curvature maxima across all scales to locate salient curve features. Mokhtarian and Mackworth [89] also show a zero crossing based curvature scale space to be an effective representation, retaining the invariance, uniqueness and stability properties. The invariance property states that two identical shapes should have the same representation, the uniqueness property states that two different shapes should have different representations and the stability property states that the variance in a given shape should result in a corresponding variance in its representation [89]. They also note practical factors which zero-crossing based multi-scale representations retain such as efficiency (the complexity of calculating and storing the representation), the ease of implementation and computation of shape properties. The latter is of importance because it addresses the part-whole problem [89,144]. Here problems arise at the boundaries of isolated contour segments, or open curves as the smoothing filter used to derive the curvature scale space will meet the edge of the contour data [89]. This problem is compounded when the blurring filter is wider than the actual contour segment. Figure 2.13 shows the curvature scale space representation for the contour illustrated in figure 2.12(a).



Dominant curve points and curvature function at scale $\sigma = 2^1$.



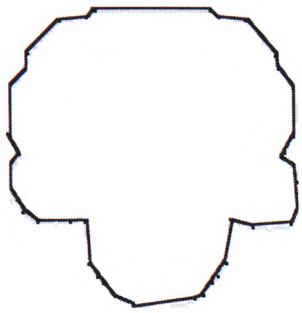
Dominant curve points and curvature function at scale $\sigma = 2^2$.



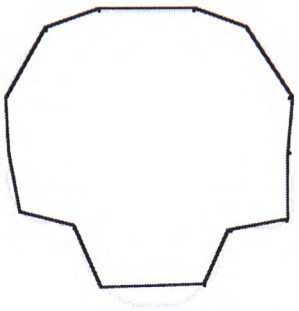
Dominant curve points and curvature function at scale $\sigma = 2^3$.

Fig 2.13. Curvature scale space representation created for the shape in figure 2.12(a). Dominant curve points are shown along with the corresponding curvature function at each scale.

Polygonal representations are also used to build shape models from the contour data in the image [84]. This involves the derivation of vertices from the boundary data using various dominant curve point detection schemes. The boundary can be approximated to different degrees of detail, eliminating small scale noise-based structures, as illustrated in figure 2.14.



Polygonal representation with a high degree of detail.



Polygonal representation with coarse detail - only salient dominant curve points are selected as vertices.

Fig 2.14. Polygonal representations with varying degrees of detail can be created by linking dominant curve points which become vertices on the polygon.

Geometrical invariance is also an important area of machine vision. It involves the ability to retain properties of an object under geometrical transforms [122]. For example, viewing a cube from a different perspective changes the visual image of the cube, but its properties remain unchanged. For this work, we are looking at 2D image data with moderate depth based rotation, so three dimensional transforms are not considered relevant here.

2.6. Deformable Contours

Given the natural variation of structures in biological image data, it is important to select appropriate representations which encode this deformation. Skeletal representations provide a general model of an region's shape by thinning the region using methods such as the medial axis transform [127]. However, the complexity of the images being analysed precludes this approach unless a suitable segmentation can be obtained. Active contour models, or snakes have been widely used for modelling natural forms in both medical imaging and biological applications [61]. Originally proposed by Kass *et al.* [56,57], snakes are energy minimising splines which fit to objects in the image. A snake utilises internal and external constraints as well as image forces to minimise an energy functional and fit the spline to the edges in the image data. Equation (5) defines the energy function which is to be minimised for the snake parametrically defined by $\mathbf{v}(s)=(x(s),y(s))$, and figure 2.15 shows how a snake tends from its

initial position to the final shape.

$$E_{snake} = \int_0^1 E_{int}(\mathbf{v}(s)) + E_{image}(\mathbf{v}(s)) + E_{con}(\mathbf{v}(s))\, ds \qquad (5)$$

Snakes alone do not solve the problem of object localisation. It is important that the snake be suitably initialised, as shown in figure 2.15(a). This can be done manually or by using an appropriate function to initialise the snake. For example, Kreho *et al.* [61] use snakes to extract the contour of dolphin fins for subsequent identification. The snake is initialised from the LoG smoothed image of the fin.

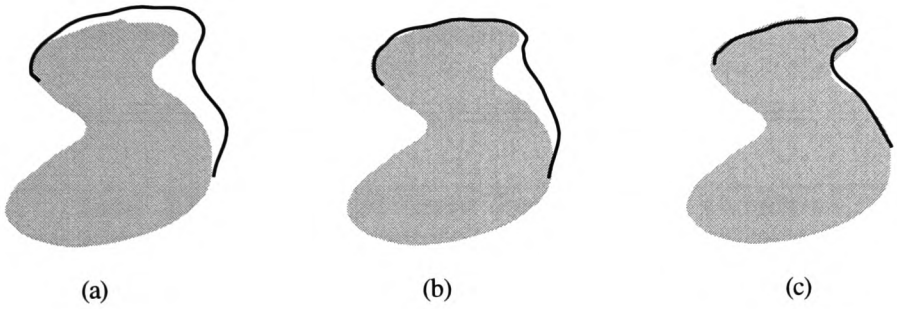


Fig 2.15. Snake optimisation. (a) Initial configuration. (b) intermediate configuration. (c) Final location of snake.

Snakes take into account the properties of the spline curve but not object shape. Lai and Chin present a generalised snake model (g-snake) which takes into account local shape curvature within a 2D shape matrix [64,65]. They study the problem of extracting objects of unknown deformation and show that the minimisation of the g-snake energy functional, which takes into account shape data, produces better object localisation results over the conventional snake, even under conditions of severe deformation. As before, initialisation of the snake is important. Lai and Chin use the generalised Hough Transform to position the snake prior to optimisation. They refer to the work of Grimson *et al.* [43] which shows that severe deformation and clutter can introduce false peaks into the Hough space (described below), resulting in sub-optimal snake initialisation. The optimisation of the snake is also affected by clutter [65]. Hence poor object localisation results are likely for cluttered, noisy images, an example of which is shown in figure 2.16.

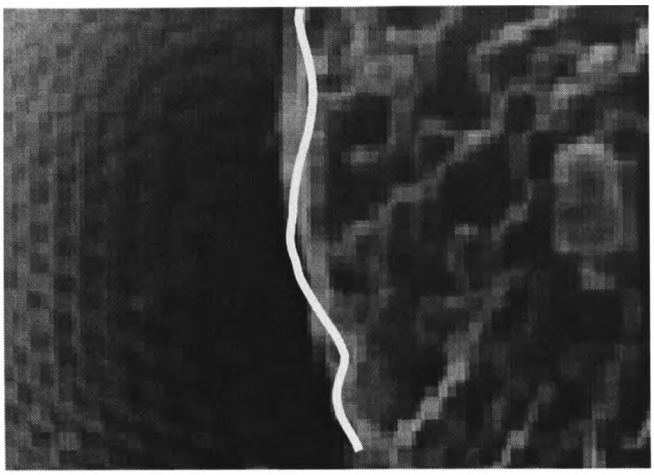


Fig 2.16. Image clutter, such as the micro-edges that define textural features, can result in sub-optimal error minimisation of the snake model and thus result in erroneous contours.

Table 2.3 summarises the shape representation techniques discussed above.

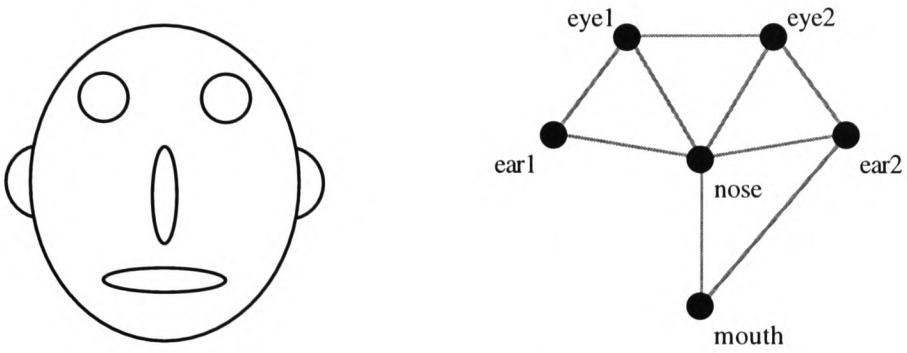
Table 2.3. Summary of Shape Representation Techniques

Technique	Useful	Reason
Syntactic Description	No	Only suited to rigid shapes, not natural deformable objects
Eigenshape Representation	No	Does not sufficiently encode the deformation properties of an object
Curvature Scale Space	Yes	Allows the general properties of a contour to be extracted and efficiently represented as a set of dominant curve points.
Deformable Contours (Snakes)	No	Local textural clutter gives rise to poor localisation and optimisation of the snake.

2.7. Shape Recognition

Once image features have been extracted and suitable intermediate representations derived, high level processes then make sense of this data in order to carry out the main function of the image analysis system, be it robot control or object classification [112]. The basis of object classification is the high level representations used to model objects of interest [122]. Throughout the literature, numerous techniques are used represent an object's shape, from grammars, or formal languages describing shape structure, predicate logic, Minsky's knowledge frames and semantic nets [122]. Statistical approaches look at delineating a multi-dimensional feature space, assigning a specific object class to each feature sub space [8,40]. Weeks *et al.* utilise the statistical properties of well-framed images of wings which form complete feature vectors [145], although wing features do not necessarily contain sufficient discriminatory information about different species. Neural networks, which are also based on statistical theory, work well for variable and noisy input data. The various architectures are described below.

As noted above, it is important any representation used be capable of storing the natural deformation inherent in the object data. At the simplest level, fixed template matching can be used to fit the input data to pre-stored objects of interest. Such techniques work well when the problem domain consists of rigid objects, but the classification performance degrades as object deformation is introduced [65]. Correlation is also used to match stored models to input image data [112]. Although techniques have been developed which incorporate invariance to global transformations such as rotation and translation, convolution remains sensitive to local deformations of the object [105]. Formal descriptions also do not lend themselves well to shapes which exhibit natural variation [5]. Elastic graph matching is an optimisation technique which finds sub graph isomorphisms between a stored model and an input image [122,149]. Nodes in the graph represent objects, or features of interest while the links between nodes represent the spatial relationships between image features, as illustrated in figure 2.17. Hierarchical graphs allow sub-graph isomorphisms to be established before evaluating higher level graphs [122]. For example, in verifying the presence of a human face, sub-graphs can represent facial features such as the eyes, nose and mouth while a higher level graph represents the whole face, or the relationships between the facial features. Graph matching has been widely used throughout the literature, and successful results can be obtained even when severe clutter is present in the input images [149].



(a) Face represented by component objects : eyes, mouth, nose and ears

(b) Graph representation of the face model

Fig 2.17. Example graph representation. (a). Face model comprised of eye, nose, mouth and ear features. (b) Graph representation of the face model, with objects represented as nodes and the spatial relationships represented as links between the nodes.

Numerous techniques have emerged based on the Hough transform [67]. Initially developed to localise analytical structures such as lines and circles, the transform maps image data to a discrete parameter space, or accumulator array, from where the location of objects can be mapped back to the image [47]. An extension of the Hough transform, the generalised Hough transform (GHT), was developed to localise arbitrary shapes which cannot be represented analytically [7]. This utilises a representation called an Rtable which indexes edge points based on their orientation. The basic accumulator array has the same dimensionality of the image but this can be augmented for rotation and scale invariance. The GHT does not scale well for more complex image data that contains noise and irrelevant image clutter since Hough's original technique was intended to extract shapes from simple image classes. Grimson *et al.* [43] show that erroneous maxima (false object locations) are as likely to be produced in the GHT as maxima at true object locations. The problem is compounded when deformation is present. Lai and Chin use the GHT to approximate the location of deformable objects within images containing moderate clutter and noise [64]. They model object deformation using a set of training instances defined with a fixed set of boundary points. Grimson *et al.* recommend the GHT be constrained to simple tasks where the above problems of noise and deformation are unlikely to arise. They also recommend that post-processing should be applied to select the maxima that correspond to true object locations. Typically the update function for the accumulator array is a "top hat" function confined to a single accumulator array element, or pixel. Smooth update functions which modify surrounding accumulator array elements based on their distance from the selected element have been proposed to overcome the problems inherent in the "top hat" update function [92]. Improved results have been reported using this approach, as shown in figure 2.18, and optimal results can be obtained in the presence of random noise [93]. Statistically, the Hough approach is equivalent to the maximum likelihood

function and is shown to produce robust results [126]. However, for more complex image data, correlated noise, or extraneous image features can give rise to false maxima in the parameter space. The use of high level knowledge can be used to segmentation of the Hough parameter space, focusing subsequent search for maxima corresponding to the true object location [68].

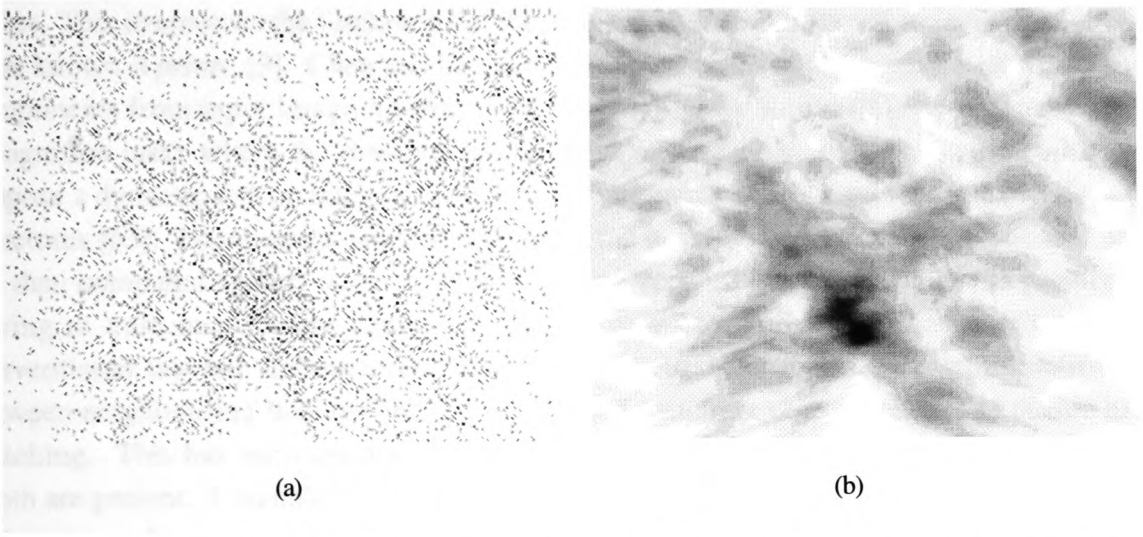


Fig 2.18. (a) Object localisation with the conventional "top hat" Hough update function. (b) Localisation done with the smooth update function. Here the location of the object is more apparent.

2.8. Neural Networks

Artificial neural networks are biologically inspired statistical tools whose properties allow information to be organised for the classification and generalisation of input patterns within a predetermined domain. They have been used in texture segmentation, image feature extraction and subsequent pattern recognition because of their ability to handle noisy data and to normalise features at the extraction phase which reduces the combinatorial explosion that can result from geometric transformations. Their inherent parallelism and ability to incorporate higher level knowledge is also considered important for classification tasks [32]. Numerous architectures incorporating these ideas have been developed. Lin and Wang [69] proposed a multi-layered architecture consisting of a position normalisation layer, a rotation normalisation layer and a feature extraction layer, the output of which is normalised for scale invariance. A similar architecture has been proposed where nodes work collectively performing image processing and feature extraction operations [5]. Widrow *et al.* [147] proposed an architecture based on multiple layers of ADALINEs (ADAptive LInear NEurons) [146], and consists of an invariance net and a trainable classifier. ART (Adaptive Resonance Theory) networks have been widely used because of their stability and plasticity properties. Srinivasa and Jouaneh

[123] used the ART1 architecture for the recognition of image features, with a later extension to include a third layer, based on the principal of invariance, to reduce the number of nodes [124]. Zheng [155] used the ART architecture in a two stage, grouping and description approach to image feature extraction and segmentation. Bhandarkar *et al.* implemented an extension to Kohonen's SOM (Self Organising Map) [60], called the HSOFM (Hierarchical Self Organising Feature Map). This was developed to build abstraction trees of images and delineate each region. This overcomes the limitations of the SOM in that the number of regions does not have to be known *a priori* [9]. Chen and Liu also use unsupervised methods to extract the principal components from input image features [17]. This addresses an important problem for object recognition tasks where the dimensionality of the input feature space is high. Lades *et al.* propose a dynamic link architecture which builds representations based on syntactically linked structures [63]. Local image features are combined to form composite feature vectors, which are then compared against other composite feature vectors. This overcomes the problems of having to train new feature detector nodes for new object classes, a problem inherent in conventional layered architectures. Links in the neural net represent spatial relationships between neighbouring features and the problem of matching objects reduces to elastic graph matching. This has been applied successfully to problems where distortion and rotation in depth are present. Carpenter *et al.* also utilise a spatial mapping neural network by extracting salient image features and feeding these into what and where channels, providing information on an objects structure and location in the image respectively [14]. Sahiner *et al.* utilise an extension of the back propagation network called the convolution neural network (CNN) [114]. The inputs to the CNN are local texture features and they show that this choice of input features is more important than the choice of architecture. This has been applied to the problem of classifying mass and normal breast tissue with a true-positive result of 90%. Tang and Stewart [133] describe a Learning Vector Quantisation (LVQ) architecture [32] which uses local textural and shape features as inputs for classifying images of plankton. They show that 95% accuracy can be obtained, highlighting the effectiveness of neural networks to biological image analysis.

2.9. Face Recognition Approaches

The problem of face recognition requires feature points to be accurately located in images of human faces. Although we are not interested in human faces, we are interested in locating landmark points on the anterior view of parasitic wasp heads. As such, the techniques used for face recognition may be relevant here. Generally, face recognition approaches fit into one of two categories. These are holistic approaches and analytical approaches [66]. Holistic methods deal with the analysis of the whole face. Here, well framed images of faces are used to derive a face space. From a set of training images, the principal components are extracted and an eigenface representation is produced [88,131,138]. Input faces are then mapped to this

face space and the most likely stored face is recalled. Such approaches are considered to be better than using facial features such as the eyes and mouth as they do not offer sufficient discriminatory power and a good initial guess of their position is required [94]. Holistic approaches suffer from the drawback that variance in the background can offer more towards face recognition than the actual facial features, with recognition performance degrading when the background was removed more so than when the actual faces were removed [87,94]. For the wasp head data, damage to specimens, as in the case of image e402, and extraneous clutter would both affect the results of this approach.

Analytical approaches look for specific structures on the head, such as the eyes, nose, mouth and chin. Lam and Yan [66] present an analytical approach where edge detection is used to find the chin feature. From there, the mouth is approximated and located accurately with a deformable snake contour. The nose and the eyes are approximated and localised in the same way. Once found, corner points are found using Xie *et al.'s* corner detection scheme [151]. Pentland *et al.* [94] show that this approach works provided the spatial relationships between features are stable and the use of snakes to localise object boundaries requires that clutter is at a minimum, which is the case for the continuous regions that define human faces. However, these techniques do not lend themselves to the wasp head image data because of the textural detail that is present. Gabor wavelets have also been used for face recognition by modelling local image properties within a conjoint spatial - spectral representation, similar to that used for texture analysis [26]. Intrator *et al.* [52] use the principal component features from a set of well framed training images as the input to a neural network for face recognition. This eliminates the problems of high dimensionality and improves recognition results. However, they map the mouth and eye features to standard locations, and as with Pentland *et al.'s* work described above, requires the spatial relationships between these features to be stable. Since the inter-species spatial relationships between the wasp head landmark points is not stable, these techniques are unsuitable for taxonomic landmark localisation. Table 2.4 summarises the shape recognition techniques discussed above.

Table 2.4. Summary of Shape Recognition Techniques		
Technique	**Useful**	**Reason**
Correlation	No	Not robust to local deformations which are prevalent in the wasp head images
Elastic Graph Matching	Yes	Allows deformable structures to be localised in an input image
Hough Transform	No	Wasp head sub-structures cannot be described analytically
Generalised Hough Transform	Yes	Non-analytical shapes can be described and localised in an input image
Holistic (Global Feature) Face Recognition	No	Specimen damage and background clutter can seriously effect the results
Analytical (Local Feature) Face Recognition	No	Spatial relationships on wasp heads are not stable like they are on human heads. This stability is a prerequisite for this technique to work
Artificial Neural Networks	Yes	Good results for species identification problems

2.10. Summary

From the research described above, wavelet based texture segmentation and multi-scale edge detection techniques are shown to be important for image feature extraction and will be investigated further. Specifically the Gabor and Mallat wavelets will be studied further. Global image processing techniques such as Fourier analysis are not considered relevant because the landmark points are local image features which must be retained. Curvature scale space will be investigated further as an intermediate shape representation. Deformable contour models are restricted and will not be investigated further because they are sub-optimal under conditions of severe image clutter. Given that the image contains significant textural detail, such approaches are likely to fail. Of the high level representation and processing techniques researched, both the generalised Hough transform and elastic graph matching approaches will be investigated further. Artificial neural networks will then be investigated for the final species identification process. These techniques are summarised in table 2.5 below.

Table 2.5. Summary of Techniques for Further Investigation		
Technique	**Useful**	**Reason**
For image feature extraction		
Wavelets (Gabor)	Yes	Retains local properties and provides good natural texture segmentation
First Derivative Edge Detectors-Mallat Wavelet	Yes	Allows important edge features to be distinguished from unimportant ones
For intermediate shape representation		
Curvature Scale Space	Yes	Allows the general properties of a contour to be extracted and efficiently represented as a set of dominant curve points.
For object localisation		
Elastic Graph Matching	Yes	Allows deformable structures to be localised in an input image
Generalised Hough Transform	Yes	Non-analytical shapes can be described and localised in an input image
For species identification		
Artificial Neural Networks	Yes	Good results for species identification problems

Chapter 3

Image Feature Extraction

3.1. Introduction

The nature of the wasp head images under consideration requires that the feature extraction techniques utilised satisfy two important criteria. First, since the landmarks of interest are localised features, it is important that local image features are retained. Secondly, the textural content of the images requires that a suitable segmentation of the image is obtained, where the boundaries of each region are accurately localised. Wavelet theory provides useful tools that retain local image properties and produce a good segmentation of images into their component regions. To determine the effectiveness of wavelets for the extraction of wasp head image features, this chapter investigates techniques for both texture segmentation and multi-scale edge detection. The Gabor wavelet will be investigated since it has been shown to produce a good segmentation of images using local orientation and spatial frequency properties [23]. For multi-scale edge detection, the Mallat wavelet edge detector will be investigated, first because it is optimal in terms of Canny's detection, localisation and one response criteria and second because it is a first derivative operator whose output allows us to discriminate between important and unimportant image features [79]. Texture analysis techniques are shown to be unsuitable because they have poor edge localisation and introduce correlated noise while multi-scale edge detection techniques eliminate correlated noise and allow the inherent displacement of edges to be eliminated. The Mallat wavelet transform is therefore used for the initial extraction of features from the wasp head images shown in appendix 1. No original contribution to knowledge is presented in this chapter as it investigates the application of low level image feature extraction techniques to the wasp head data shown in appendix 1.

3.2. Texture Region Segmentation

To demonstrate the effectiveness of Gabor based texture segmentation on the SEM images in appendix 1, a bank of Gabor wavelets configured over the orientation range $[0,2\pi)$ and spatial frequency range [0.75,1.5] was set up. Each filter was convolved with test images from appendix 1 and the resulting filter responses formed a feature vector describing each point in the image. This is not an optimised approach to texture segmentation, where the number of filters are reduced based on the principal features of the texture. However, it does illustrate the effectiveness of texture segmentation applied to the wasp head images used in this work. From this feature vector, Porat and Zeevi's orientation, frequency and intensity criteria were used to segment each textural region in the image [99]. The mean and variance of the texture orientation at each point are calculated using equations (6) and (8) respectively,

$$\hat{\theta} = \frac{1}{\sum_{\theta}\sum_{f} |M_{\theta f}|} \sum_{\theta}\sum_{f} \theta \; |M_{\theta f}| \tag{6}$$

$$\delta\theta_{\theta f} = |\hat{\theta} - \theta| \qquad \text{(for all } \theta,f) \tag{7}$$

$$sd\theta = \frac{1}{\sum_{\theta}\sum_{f} |M_{\theta f}|} \sum_{\theta}\sum_{f} \delta\theta_{\theta f} \; |M_{\theta f}| \tag{8}$$

where $\delta\theta_{\theta f}$ represents the variance of each orientation filter response from the mean orientation. Similarly, the mean and variance of the texture frequency can be calculated from the Gabor filter responses. These are defined in equations (9) and (11) respectively where $M_{\theta f}$ represents the filter response at orientation θ and frequency f.

$$\hat{f} = \frac{1}{\sum_{\theta}\sum_{f} |M_{\theta f}|} \sum_{\theta}\sum_{f} f \; |M_{\theta f}| \tag{9}$$

$$\delta f_{\theta f} = |\hat{f} - f| \qquad \text{(for all } \theta,f) \tag{10}$$

$$sd\,f = \frac{1}{\sum_{\theta}\sum_{f} |M_{\theta f}|} \sum_{\theta}\sum_{f} \delta f_{\theta f} \; |M_{\theta f}| \tag{11}$$

Figure 3.1(a) shows the resulting spatial frequency feature image and figure 3.1(b) shows the resulting orientation image. Combining these features with the local image intensity, a feature vector is created at each point in the image. This is then used to segment the image into its component regions. Figure 3.1(c) illustrates the constituent regions that are identified while figure 3.1(d) shows the final edge map for the identified regions. Two important problems can be identified from these results. The first major problem is the emergence of spurious regions, or correlated noise. This results from the natural variation present in the texture within a given image region. It is acknowledged throughout the literature that such variance in the properties of the texture can give rise to erroneous regions. This problem also occurs for optimised texture filtering approaches [102]. The second major problem is the displacement of region boundaries. Throughout the literature, results have shown that the localisation of boundaries between textural regions is a difficult problem, as the properties of neighbouring regions

converge in the area around each edge point. Randen and Husøy attempt to overcome these problems by modelling the edges between textural regions [102], while Ganesan and Bhattacharyya [38] attempt to directly extract the edges between textural regions. Their results show poor localisation of edge points for textured images. Davis and Mitiche [25] also consider directly extracting edges between textured regions, but they only consider simple, structured textures. Since landmark points reside on the edges of important image structures, as shown in appendix 2, the displacement of edges inherent in texture segmentation approaches can affect the accuracy of the final landmark locations and hence affect the resulting species classification.

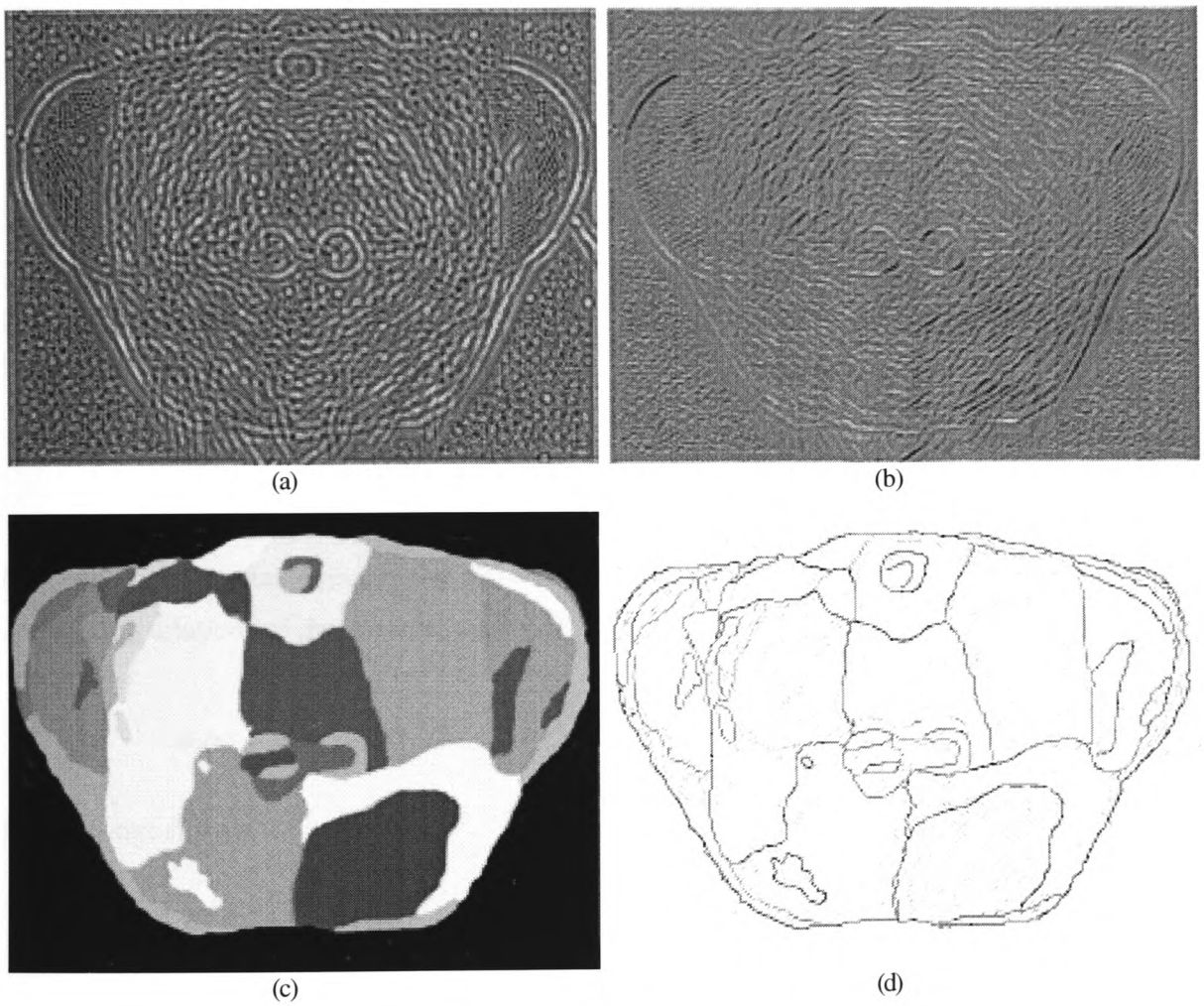


Fig 3.1. (a) Spatial frequency feature image. (b) Orientation feature image. (c) Resulting segmentation of the test image. (d) Final contour map showing the edges between each identified region.

3.3. Multi-scale Edge Detection

Given the problems of poor edge localisation and the introduction of correlated noise inherent in texture segmentation approaches, it is important to consider other image feature extraction techniques. Since multi-scale edge representations lend themselves to the extraction of salient contour information and the filtering of irrelevant textural clutter, it is important to investigate their applicability to the wasp head images in appendix 1. Since first derivative filter responses have advantages over second derivative filters in that important image features can be distinguished from noise-based features [79], this work work will focus on the Mallat wavelet [79]. Since this is a scalable operator which approximates the first derivative of the gaussian smoothing function, it allows multi-scale representations which retain the causality and semi-group properties [150] to be created.

3.3.1. Mallat Wavelet Decomposition

The Mallat wavelet is defined as a quadratic spline approximation of the first derivative of the Gaussian smoothing function $\theta(x)$ which is sufficiently localised in time and frequency [19,79].

$$\psi(x) = \frac{d\theta(x)}{dx} \tag{12}$$

For one dimensional signals, the wavelet transform at scale s, $W_s f(x)$, is obtained by convolving dilations of the "mother wavelet" $\psi_s(x)$ where

$$\psi_s(x) = \frac{1}{s}\,\psi\left(\frac{x}{s}\right) \tag{13}$$

with the signal $f(x)$.

$$W_s f(x) = f * \psi_s(x) \tag{14}$$

The one dimensional wavelet transform can be extended into two dimensions using two orthogonal wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ approximating the first derivative of the two dimensional Gaussian smoothing function $\theta(x,y)$.

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \qquad \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y} \tag{15}$$

For two dimensional images, the wavelet transform at scale s contains two components, W_s^1 and W_s^2, obtained by convolving the wavelet dilations $\psi_s^1(x)$ and $\psi_s^2(x)$ where

$$\psi_s^1(x,y) = \frac{1}{s^2}\psi^1\left(\frac{x}{s},\frac{y}{s}\right) \qquad \psi_s^2(x,y) = \frac{1}{s^2}\psi^2\left(\frac{x}{s},\frac{y}{s}\right) \tag{16}$$

with the image f(x,y).

$$W_s^1 f(x,y) = f * \psi_s^1(x,y) \qquad W_s^2 f(x,y) = f * \psi_s^2(x,y) \tag{17}$$

The gradient vector $M_s f(x,y)$ indicating the direction of the combined wavelet response in the image plane is defined as

$$M_s f(x,y) = \sqrt{|W_s^1 f(x,y)|^2 + |W_s^2 f(x,y)|^2} \tag{18}$$

and the orientation of the gradient vector $A_s f(x,y)$ is given by

$$A_s f(x,y) = \tan^{-1}\left(\frac{W_s^2 f(x,y)}{W_s^1 f(x,y)}\right) \qquad \text{where } 0 \le A_s f(x,y) < 2\pi. \tag{19}$$

For one dimensional signals, maxima in the modulus of the wavelet response correspond to sharp gradient changes. In two dimensions, local maxima along the gradient of the combined wavelet response correspond to individual edge points. This is equivalent to the output of Canny's algorithm and hence optimal in terms of Canny's detection, localisation and one response criteria [79]. The behaviour of maxima across each scale in the wavelet transform contains important information about contours in the image. Mallat and Zhong show that maxima decay across multiple scales characterises different types of edge [79]. To investigate this further, a set of isolated one dimensional step edge functions $F_{\sigma A}(x)$ are analysed where

$$F_{\sigma A}(x) = \begin{cases} \begin{pmatrix} -A \text{ if } x < 0 \\ A \text{ if } x \ge 0 \end{pmatrix} & \text{if } \sigma = 0 \\ \dfrac{A}{1+e^{\frac{-x}{\sigma}}} & \text{otherwise} \end{cases} \tag{20}$$

and σ and A represent the continuity and amplitude of the step edge respectively. The one dimensional wavelet transform as defined in (14) is equivalent to the cross-correlation of the wavelet dilation at scale s, $\psi_s(x)$, with the step edge function $F_{\sigma A}(x)$. The wavelet response at

each scale indicates the relevance of each edge event at that scale, and only begins to decay when the gradient of the wavelet $\delta\psi_s(x)$ is greater than the gradient of the step edge $\delta F_{\sigma A}(x)$. This results in the characteristic positive decay where the maxima values increase proportionately to the scale [79]. Negative decay, where maxima values decrease as the scale increases, results when isolated edge events with positive or zero decay are analysed in the global context of the signal, or image. For each scale of analysis, *s*, negative decay results when

$$\frac{R_{s+1}(x)}{M_{s+1}} < \frac{R_s(x)}{M_s} \tag{21}$$

where $R_s(x)$ is the wavelet response for edge event x at scale s and M_s is the maximum wavelet response at scale s. The spatial extent of each feature in the image is also important in determining the decay of maxima across multiple scales. Figure 3.2(a) shows a one dimensional signal extracted from the image in figure 1.1(a), with the spatial extent of the singularity at abscissa 2 increasing monotonically.

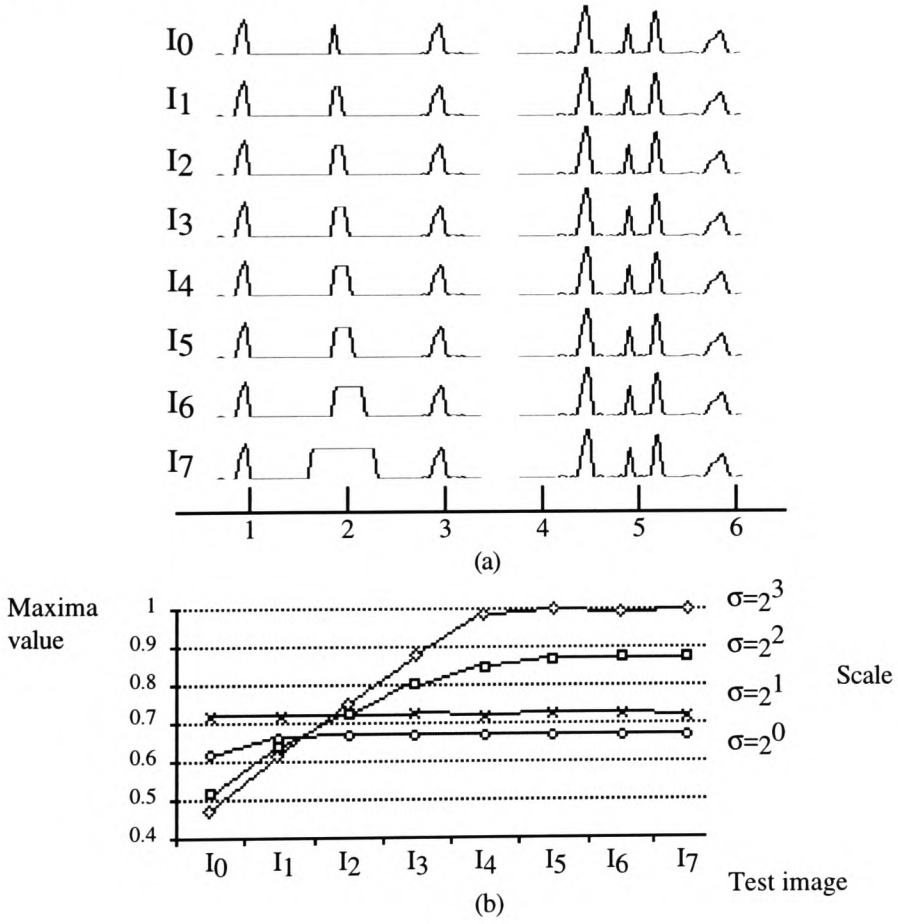


Fig 3.2. The decay of maxima that correspond to edge events across scale space is related to the spatial extent of image features.

The wavelet transform modulus maxima exhibit negative decay for the events that define this singularity, which inverts to positive decay as the spatial extent increases, as shown in figure 3.2(b). Given the cross correlation properties of the wavelet transform, maxima responses indicate the relevance of edge events at each scale. Given that maxima exhibit negative decay only in the global context of each signal, the maxima at the largest scale of support, or the scale at which a given edge event emerges in a coarse to fine scale traversal, is the most relevant.

In creating a multiscale representation, it is important that sufficient information about the original image is retained, although for object recognition problems this is less critical than for signal reconstruction [75]. Hence the choice of scale range is important. It is well known in information theory that logarithmic sampling leads to informationally uniform sampling density [33]. Therefore a dyadic scale sequence is used to retain the salient information in an image, even though it is over-complete, retaining redundant information [79]. Discrete images have a limited spatial scale range, bounded by the sampling resolution and the spatial extent of the features within the image [59]. For the SEM images in appendix 1, salient landmark information is contained within the dyadic scale range $\sigma=_2 0$ to $\sigma=_2 3$, and throughout this work, the Mallat wavelet transform and subsequent analysis is confined to this range. Figure 3.3 shows the multi-scale representation created for image e100 in appendix 1.

3.3.2. Localising Displaced Maxima

In order to determine the characteristics of wavelet responses across multiple scales it is important to find correspondences between maxima at each scale of analysis. In creating a multiscale representation, maxima at larger scales do not occur at their true location [33,150]. This displacement results from the blurring required to eliminate smaller scale features and is determined by the width of the underlying Gaussian smoothing function and the interaction with surrounding edge events [139]. For one dimensional signals, individual points are displaced creating characteristic scale space arches [150]. In two dimensions, individual points become contours which deform, split and merge as the scale of analysis increases [33,70,75]. The problem becomes more tractable when an image is analysed across a sufficiently dense range of scales so the displacement of edge points is bounded by 1 pixel [75,79]. However, this is computationally expensive and for this work it is important to link wavelet edge events across a dyadic scale range (a logarithmic scale range of base 2).

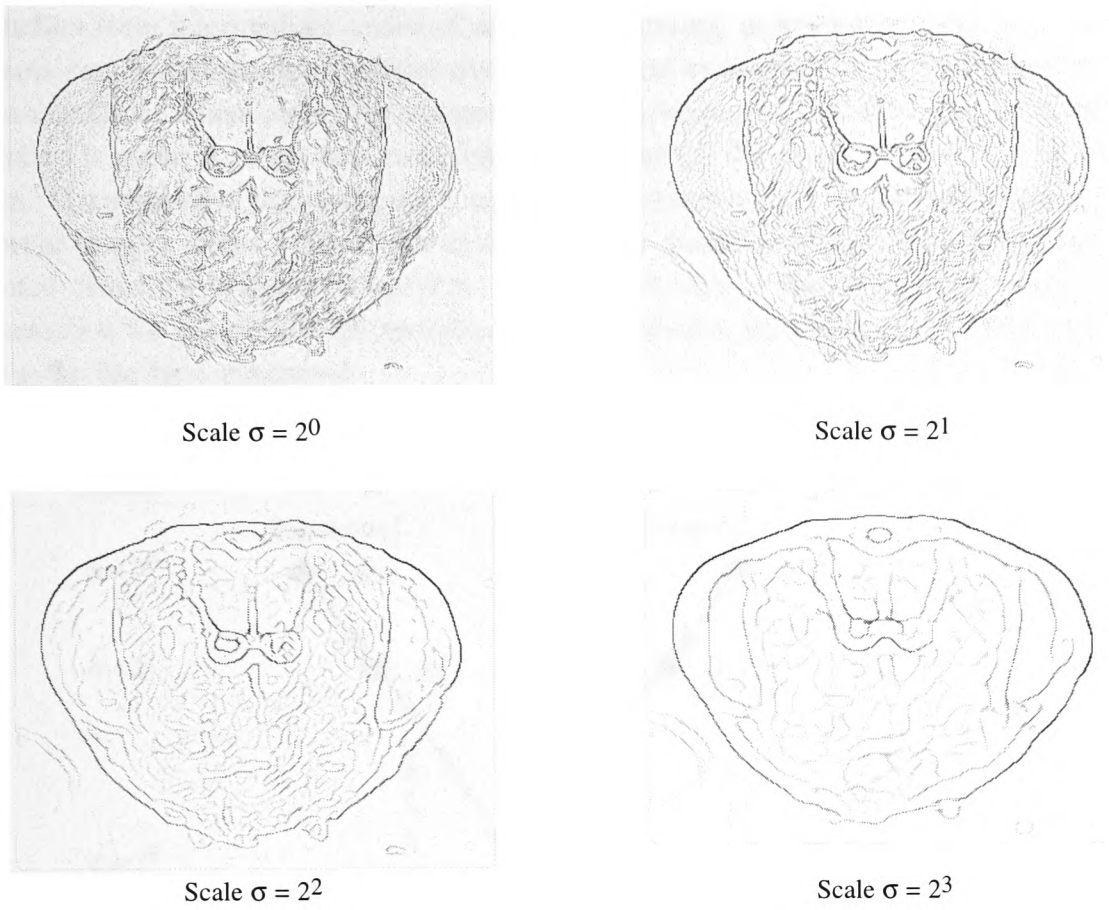


Fig 3.3. Wavelet transform for image e100 in appendix 1. Edges split, merge and change location as the scale of analysis increases.

Lu and Jain's RESS (Reasoning about Edges in Scale Space) system [75] employs the CEL (Correcting Edge Location) procedure which finds correspondences between individual edge points at adjacent scales. This procedure looks for corresponding zero crossings at adjacent scale levels within a circular region of interest bounded by the Gaussian width σ. Correspondences are found between zero crossings and their closest neighbours at adjacent scales, that is those zero-crossings with the smallest radial distance. Here a similar approach is adopted to find correspondences between maxima in the Mallat wavelet transform. However, neighbouring edge points on the same contour should be ignored and only maxima along the direction of the wavelet response are considered. Correspondences are found according to the relative radial and angular positions as well as the intensity of the maxima. Due to the splitting and merging of contour points that occurs in two dimensional scale space and the emergence of false edges at larger scales, there is no guarantee that a 1:1 correspondence exists between maxima at adjacent scales. As a result discontinuities occur in the localised contours at larger scales. Also, given the sparse analysis density around larger scales in the dyadic range

described above, errors in the correspondences found are more likely to occur since information from intermediate scales of analysis is missing, as shown in figure 3.4. These problems can be overcome by iteratively creating and evaluating correspondences found between maxima. Where more than one possible correspondence exists for a given maximum, a selection is made as to the best correspondence based on the above position and intensity criteria. The problem of discontinuous contours at larger scales can be resolved by looking for support at smaller scales. This is discussed in greater detail later when contour models are extracted from the wavelet transform. Figure 3.5 shows the final multi-scale edge representation for image e100 in appendix 1. As can be seen, the deformation of the edges at larger scales has been eliminated.

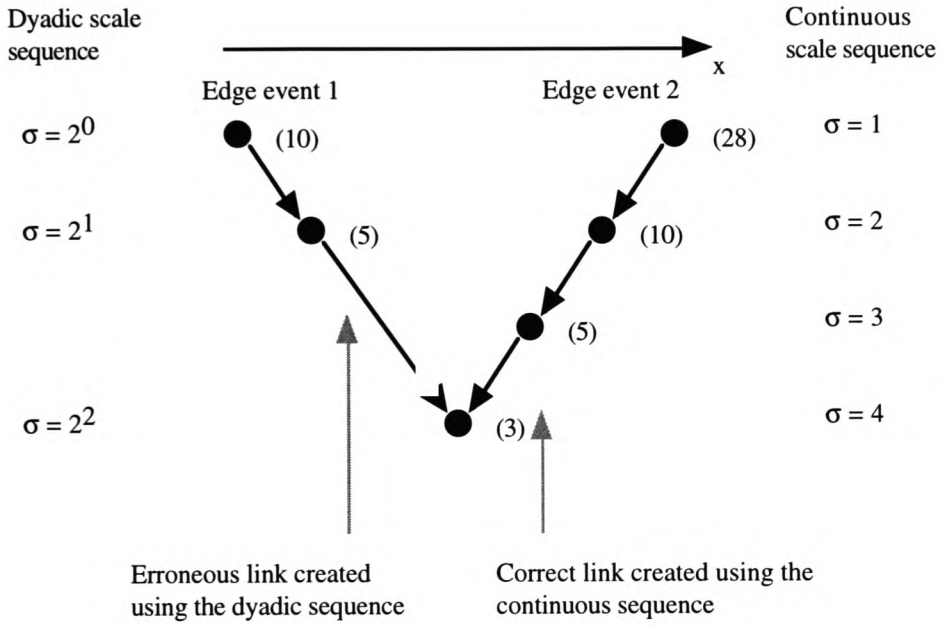


Fig 3.4. Sparse information at larger scales of analysis in the dyadic scale range can result in erroneous links between maxima. The maxima as scale $\sigma=2^2$ belongs to edge event 2. Missing information in the dyadic scale range means that an erroneous link to edge event 1 is made.

3.3.3. Identifying and Eliminating False Contours

A major problem which occurs with the splitting and merging of contour points is that false edges can emerge at larger scales of analysis [33,75,153]. This is an unavoidable effect of Gaussian based edge detection operators. One of the important properties of Gaussian derived multiscale representations is that all contours can be tracked to their true location at the smallest scale of analysis [150]. False edges can be identified when they have no support at smaller scales. Thus, once correspondences have been found between maxima, false edge points can

be eliminated by removing points which have no corresponding maximum at the adjacent smaller scale. Applying this rule in a fine to coarse scale traversal allows false edges that emerge at any given scale and perpetuate at larger scales to be eliminated.

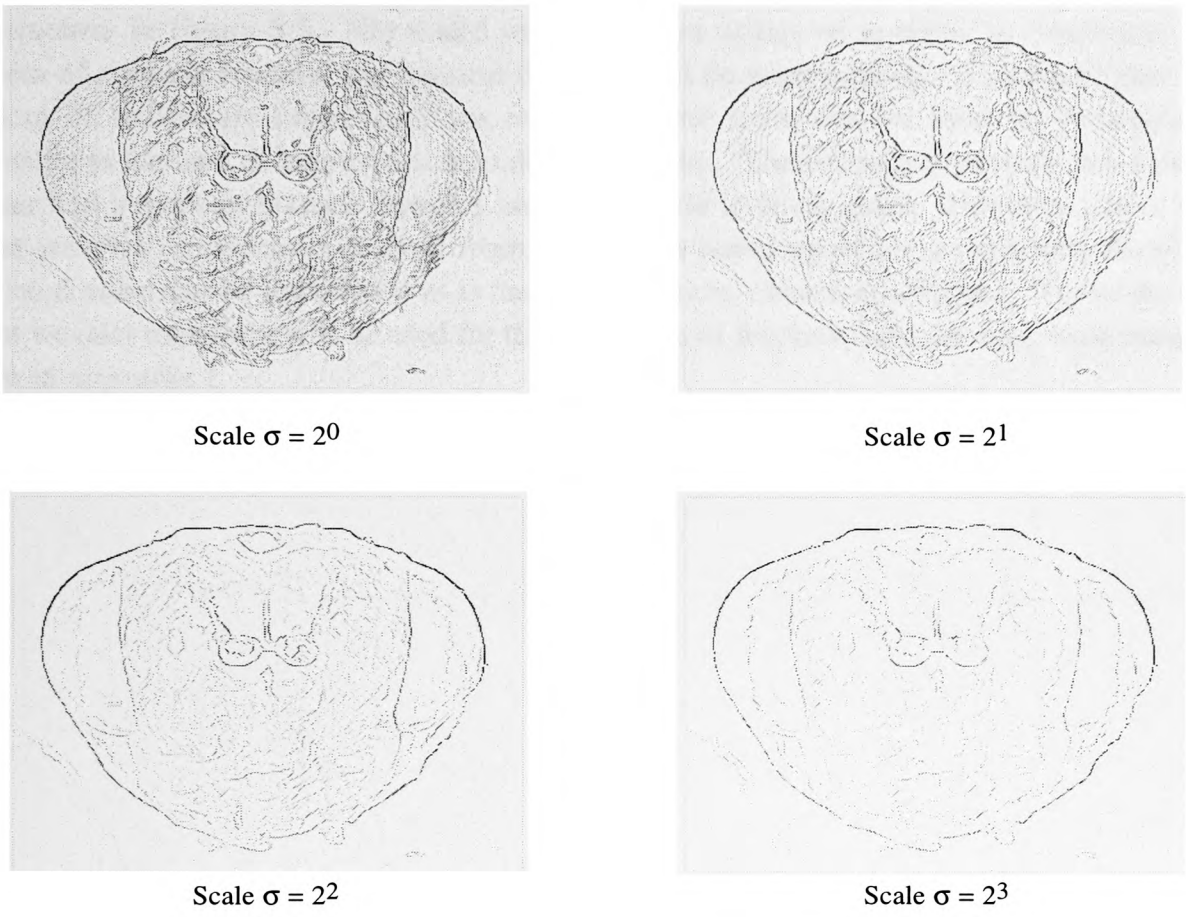


Fig 3.5. Correcting edge displacement by finding correspondences between maxima at adjacent scales of analysis. The distortion present in figure 3.3 has now been eliminated.

3.4. Conclusion

Both texture analysis and multiscale edge detection techniques have been investigated. Given the variance in the textural properties of the SEM images being studied, the use of texture analysis methods results in poor edge localisation and the introduction of correlated noise, or erroneous regions. The Mallat wavelet transform has also been studied. This too suffers from the problem of poor edge localisation, but only at larger scales of analysis. This problem can be rectified since edge events in scale space can be tracked back to their true location in the image. Comparing figure 3.5 with figure 3.3 shows that edge displacement and inherent

deformation can be eliminated. Figure 3.5 also shows that textural detail can be eliminated at larger scales of analysis (appendix 4 shows the wavelet transforms for all of the images used in this study). However, not all textural detail can be eliminated as larger scale textural features are retained while important contour features that define the wasp head sub-structures of interest begin to erode at larger scales of analysis. This can be seen at the bottom of the eye sub-structure in figure 3.5. We could utilise smaller scales of analysis to overcome the problem of contour erosion, but textural detail would be reintroduced. This highlights the problem of multiscale representations where no one scale can be used for subsequent processing as features of interest reside at different scales. These results also show that salient contour and irrelevant textural features are inseparable in scale space. However, since the Mallat wavelet transform is derived from a gaussian based operator, no spurious detail or correlated noise can be introduced as is the case for texture based approaches. Therefore the Mallat wavelet transform will be used for the extraction of features from the wasp head images shown in appendix 1.

Chapter 4

Contour / Texture Delineation

4.1. Introduction

The Mallat wavelet transform described in the previous chapter is used to derive a multiscale representation of the SEM images in appendix 1, where significant textural clutter can be filtered out at larger scales of analysis. Although textural clutter can be reduced, this approach increases the amount of data with which to work since 4 separate edge map images are created in the wavelet transform (one edge map image per scale of analysis). Selecting a single scale of analysis will not work given that larger scales erode salient contour features while smaller scales contain unwanted textural clutter. Thus it is important to derive a suitable representation for subsequent landmark localisation and species classification. Since textural clutter is a major problem, it is important to minimise this while maximising the presence of contour features that define the structures of interest on the wasp heads, as shown in appendix 3. The Mallat wavelet transform alone is not sufficient as salient contour and large scale textural features are inseparable in scale space. Filtering large scale textural features would result in significant erosion of the contour data we want to retain. To overcome this problem, a novel "texture filtering" technique is presented which recombines the Mallat wavelet transform to form a single contour map where salient edge data is retained while textural clutter is eliminated. This also eliminates the redundancy present in the wavelet transform by reducing the data available to a single edge map. This technique represents part of the main contribution to knowledge for this work.

4.2. Contour / Texture Properties

Understanding the nature of texture presents a non-trivial problem in that it is difficult to define what actually constitutes textural features. Though no formal definition of texture is attempted here, the features deemed important in discriminating between contour and textural features are investigated. The context in which the image is analysed, the wavelet transform modulus maxima and the local activity along the gradient of the wavelet response are considered important features. Context is important, and in the study of contour and textural features, a duality exists in that features can be perceived as either. The wavelet response at a given scale represents an edge point. However, the same point can also be considered part of a texture when analysing the image in the context of a larger scale. Multiscale representations, such as that created with the Mallat wavelet transform described above, allow images to be analysed within different contexts, or scales. Consider the "brick wall" image in figure 4.1(a) and the resulting Mallat wavelet transform in figure 4.1(b).

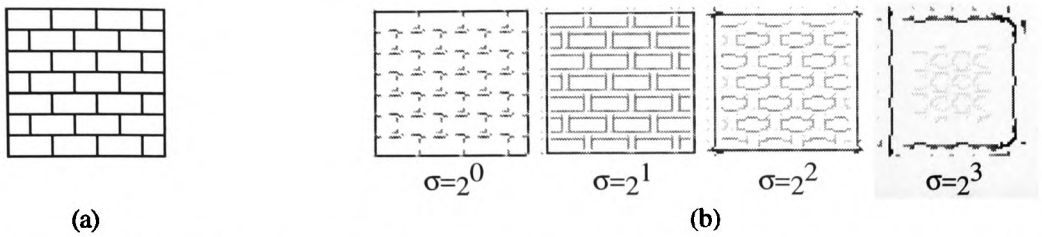


Fig 4.1. (a) "Brick Wall" Texture. (b) Wavelet transform of the texture across the scale range $\sigma=2^0$ to $\sigma=2^3$.

At smaller scales of analysis, the features defining each brick are present while in the context of the largest scale, the brick features are eliminated and the wall as a whole can be seen. Therefore, as the context changes and the scale of analysis becomes larger, so texture emerges. Hence, texture does not exist at the scale its constituent features reside, but only emerges in the context of larger scales of analysis. The image in figure 4.1(a) presents an idealised case where all of the features below the largest scale of analysis can be considered textural features. As noted above, natural scenes can contain important contour information at any scale, so considering only the largest scale of analysis may ignore important features contained in the multiscale representation. The second important feature is the set of maxima within the wavelet transform. The cross-correlation properties of the wavelet transform were investigated in chapter 3 and it was shown that the maxima at the largest scale of support, the scale at which the edge event emerges in a coarse to fine scale traversal, was most relevant. Finally the interaction with surrounding maxima, or the local activity surrounding each maximum in the wavelet transform is deemed important. If we consider the multiscale representation derived from a 1D cross-section of the eye in figure 1.1(a), shown in figure 4.2, we see that edge features exist as isolated events at their largest scale of support while textural features reside at smaller scales of analysis in densely populated regions of the wavelet transform.

To extract important edge events and filter out textural clutter a novel algorithm is presented which analyses the emergence of maxima in a coarse to fine scale traversal and combines the multiscale representation created with the Mallat wavelet transform to produce a single contour map. This uses the wavelet transform modulus maxima at the largest scale of support along with the local activity surrounding each event. This is determined by both the spatial frequency of the texture and the context in which it is analysed. Homologous regions are defined by isolated contour events with little or no surrounding activity. Depending on the context of analysis, textured regions can be bounded by texture of different spatial frequency and orientation [23] or by smooth regions. The latter results in the decay of local activity around contour points while the former presents a class of boundary which cannot easily be identified from the wavelet transform itself. Such regions require a two-phase approach where texture region segmentation is applied to the textured regions identified in the contour-texture delineation process described below. The algorithm to delineate contour and textural features is

summarised below and described in greater detail in the next section. The algorithm consists of the following steps...

Step 1: Calculate Contour Responses

For each edge event that emerges at each scale of analysis in a coarse to fine scale traversal the following steps are performed...

- Calculate ε : the square region of interest at the current scale centred on the current point (that is the wavelet maximum at the largest scale of support for the respective edge event)

- Calculate Θ : the set of neighbouring edge points along the axis perpendicular to the direction of the edge on which the current point lies.

- Calculate the interaction of each point in Θ with the current point.

- Calculate Contour Response of the current point.

Step 2 : Threshold Contour Responses at each scale

Step 3 : Logically OR all contour responses to create the final contour map

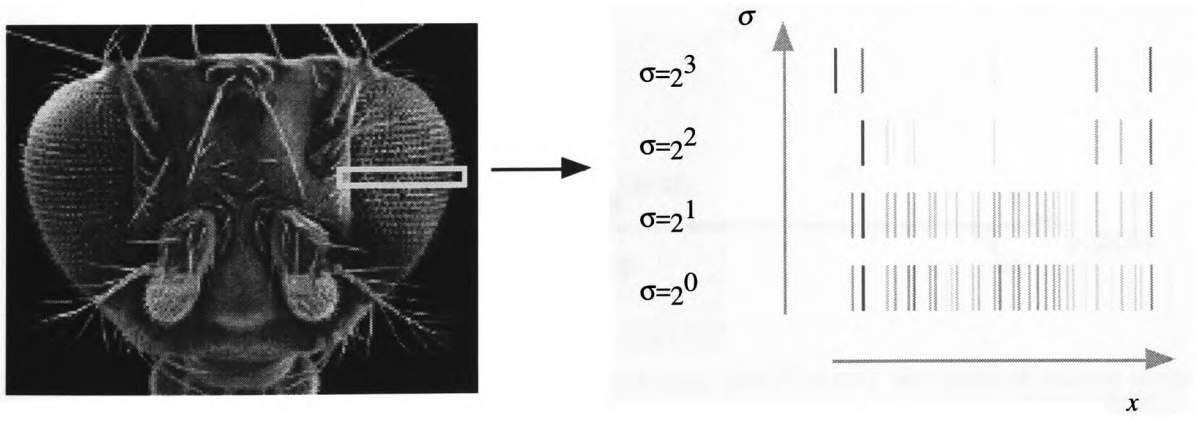


Fig 4.2. Maxima that define important edge events exist as isolated points at the high end of the scale range. Textural features exist as maxima in densely populated regions of the wavelet transform at the low end of the scale range.

4.3. Delineating Contour and Textural Features

For each maximum $M_{2^j xy}$ that emerges at scale $\sigma=2^j$, the local activity, or interaction with neighbouring events, $I_{2^j xy}$, is calculated based on the relative position and intensity of neighbouring maxima within the square region of interest $\varepsilon x \varepsilon$, centred on $M_{2^j xy}$, where ε is defined as

$$\varepsilon = \frac{1}{4} 2^{(2^{n-1} - 2^j)} \tag{22}$$

and n equals the number of scales of analysis in the initial wavelet transform. Figure 4.3 shows the resulting region of interest (ROI) curve over the scale range $[\sigma=2^0,\sigma=2^3]$ where n =4. Given this curve, the interaction between maxima at smaller scales is calculated over a larger area. For edge events at smaller scales to be recognised as contour points it is important they reside within sparsely populated regions of the wavelet transform.

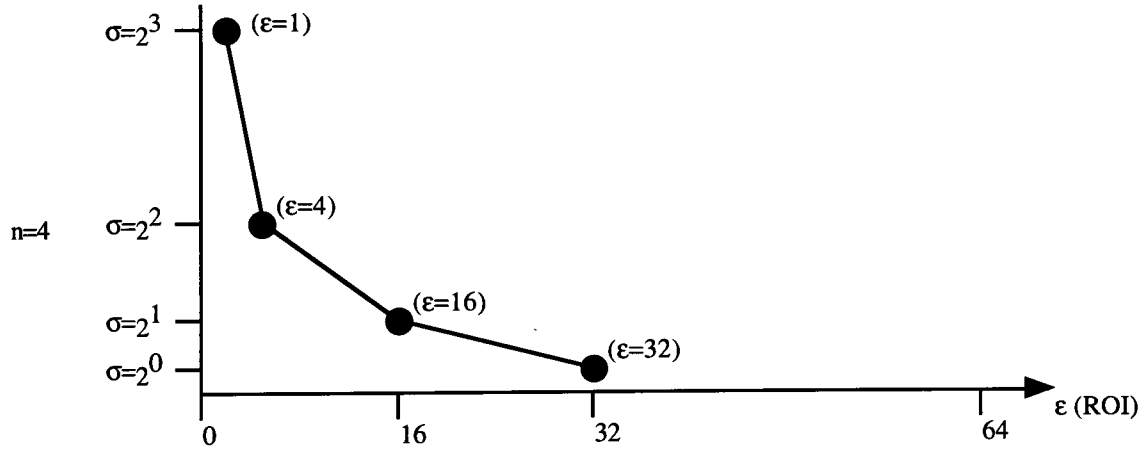


Fig 4.3. Given n=4 scales of analysis over the dyadic scale range $[\sigma=2^0,\sigma=2^3]$, the region of interest within which maxima interaction is calculated expands exponentially.

Since maxima correspond to individual edge points, neighbouring maxima on the same contour are not relevant so only those maxima that lie along the axis specified by the orientation of the gradient of the wavelet response are considered. This axis is defined by $A_{2^{j+1} xy}$, given in (19) and $A'_{2^{j+1} xy}$ which is defined as

$$A'_2{}^{j+1}{}_{xy} = \begin{cases} A_2{}^{j+1}{}_{xy} + \pi & \text{if } A_2{}^{j+1}{}_{xy} < \pi \\ A_2{}^{j+1}{}_{xy} - \pi & \text{otherwise.} \end{cases} \tag{23}$$

For each neighbouring maximum $M_2{}^j{}_{x'y'}$, where (x',y') belongs to the set of points

$$\{x',y' : x-(\varepsilon/2) \le x' < x+(\varepsilon/2) \; ; \; y-(\varepsilon/2) \le y' < y+(\varepsilon/2) \} \tag{24}$$

the radial distance $\lambda_{x'y'}$ from $M_2{}^j{}_{xy}$ is defined as

$$\lambda_{x'y'} = \sqrt{|x-x'|^2 + |y-y'|^2} \tag{25}$$

while the orientation of $M_2{}^j{}_{x'y'}$ with respect to $M_2{}^j{}_{xy}$ is defined as

$$\theta_{x'y'} = \tan^{-1}\left(\frac{y-y'}{x-x'}\right) \qquad \text{where } 0 \le \theta_{x'y'} < 2\pi. \tag{26}$$

Let Θ be the set of all relevant points $M_2{}^j{}_{x'y'}$ that satisfy the inequality

$$\delta\theta_{x'y'} < MAX\left(\frac{2\pi}{8\,(MAX(|x-x'|,|y-y'|))}, t\right) \tag{27}$$

where t is a suitable orientation threshold defining the angular window of interest with respect to the axis given in (19) and (23), and $\delta\theta_{x'y'}$ is the orientation offset of neighbouring maximum $M_2{}^j{}_{x'y'}$ relative to this axis which is defined as

$$\begin{aligned} \delta\theta_{x'y'} = MIN\,(& |\theta_{x'y'} - A_2{}^{j+1}{}_{xy}|, |\theta_{x'y'} - (A_2{}^{j+1}{}_{xy} + 2\pi)|, \\ & |\theta_{x'y'} - (A_2{}^{j+1}{}_{xy} - 2\pi)|, |\theta_{x'y'} - A'_2{}^{j+1}{}_{xy}|, \\ & |\theta_{x'y'} - (A'_2{}^{j+1}{}_{xy} + 2\pi)|, |\theta_{x'y'} - (A'_2{}^{j+1}{}_{xy} - 2\pi)|) \end{aligned} \tag{28}$$

where MAX and MIN are functions which return the argument with the largest and smallest value respectively. Given the orientation inaccuracy around neighbouring pixels inherent in square tessellated discrete images, if $t < \pi/4$ then important neighbouring pixels can be ignored. Hence the argument $2\pi/8(MAX(|x-x'|,|y-y'|))$ in (27) ensures neighbouring pixels in close proximity to $M_2{}^j{}_{xy}$ whose minimum angular distance is greater than t are considered. The

interaction $I_2{}^j{}_{xy}$ of $M_2{}^j{}_{xy}$ with the relevant neighbouring maxima is calculated with a modified convolution operation where $M_2{}^j{}_{xy}$ is ignored.

$$I_2{}^j{}_{xy} = \sum_{\substack{y'=y-\varepsilon/2 \\ (y' \neq y)}}^{y' \leq y+\varepsilon/2} \sum_{\substack{x'=x-\varepsilon/2 \\ (x' \neq x)}}^{x' \leq x+\varepsilon/2} e^{\frac{-\lambda_{x'y'}{}^2}{4\varepsilon}} M_2{}^j{}_{x'y'} \qquad \text{where } (x',y') \in \Theta \tag{29}$$

The contour response $C_2{}^j{}_{xy}$ for each edge event $M_2{}^j{}_{xy}$ that emerges at scale $\sigma=2^j$ combines the interaction of the event with the wavelet transform modulus maximum and is defined as

$$C_2{}^j{}_{xy} = \frac{1}{e^{(I_2{}^j{}_{xy} / MI_2{}^j)\, c}} M_2{}^j{}_{xy} \qquad \text{where } 0 \leq C_2{}^j{}_{xy} \leq 1 \tag{30}$$

where $MI_2{}^j$ is the largest interaction value for maxima that emerge at scale $\sigma=2^j$ and c is a scaling factor which determines the relevance of the interaction with neighbouring events. When $c \cong 4$, isolated events produce a large contour response while non-isolated events produce a small contour response. The contour responses at each scale are then thresholded to eliminate non-salient features and retain only those contour events that emerge at the respective scale. The thresholded contour responses at each scale are then logically ORed to produce the final contour map.

4.4. Experimental Results

The algorithm described above has been applied to electron-microscopic images of biological specimens and synthetic images to test its effectiveness in retaining salient contour features while filtering out textural clutter. For each image, the Mallat wavelet transform is calculated, the displacement of maxima at larger scales is resolved and false edges which emerge are removed. For each scale of analysis, the contour response is calculated using the ROI curve defined in (22) with a scaling factor c =4 and then thresholded using a value of 0.125. This threshold has been found through experimentation, applying the above algorithm to different synthetic and SEM images. The results are then combined to form a complete contour map of the image.

The first experiment analyses the electron-microscopic image of the head of an adult diptera (Fruitfly) shown in figure 1.1(a). This contains important contour information across all scales of analysis as well as textured regions. The localised wavelet transform maxima are shown in

figure 4.4 while the contour responses for each of these images are shown in figure 4.5. Combining each contour response produces the final contour map shown in figure 4.6. The results show that the algorithm effectively eliminates regions of texture while retaining important contour events, including those that reside at smaller scales of analysis. Discontinuities or weak contour responses result for edge events that define features of small spatial extent such as the elongated structures present in figure 1.1(a). Although no activity exists around these features, two edge events in close proximity results in high interaction, making it difficult to distinguish between maxima of large intensity which are in close proximity and a textural region of weak intensity.

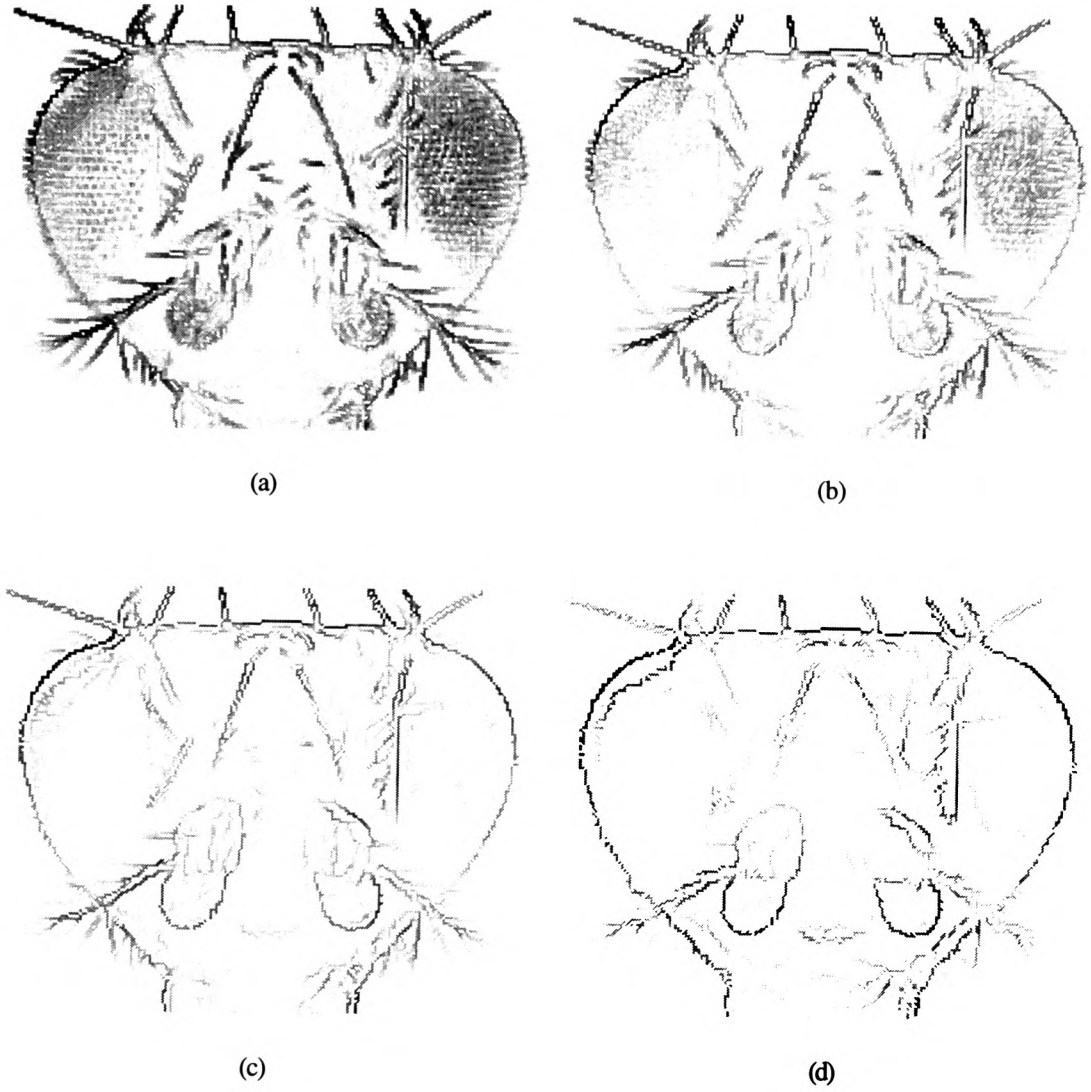


Fig 4.4. Localised wavelet transform modulus maxima for the image in figure 1.1(a) at scales (a) $\sigma=2^0$, (b) $\sigma=2^1$, (c) $\sigma=2^2$ and (d) $\sigma=2^3$.

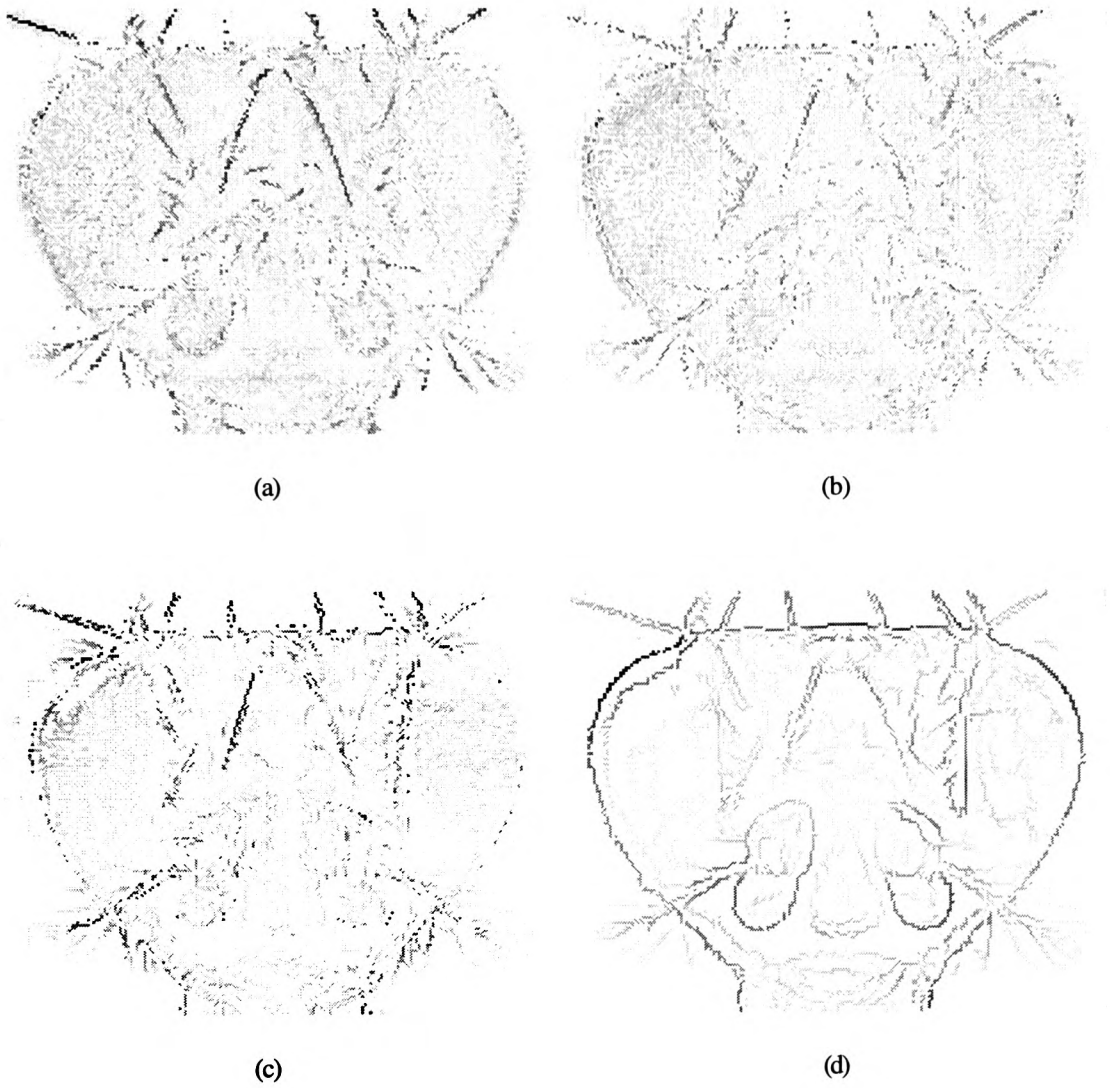


Fig 4.5. Contour responses for the image in figure 1.1(a) at scales (a) $\sigma=2^0$, (b) $\sigma=2^1$, (c) $\sigma=2^2$ and (d) $\sigma=2^3$.

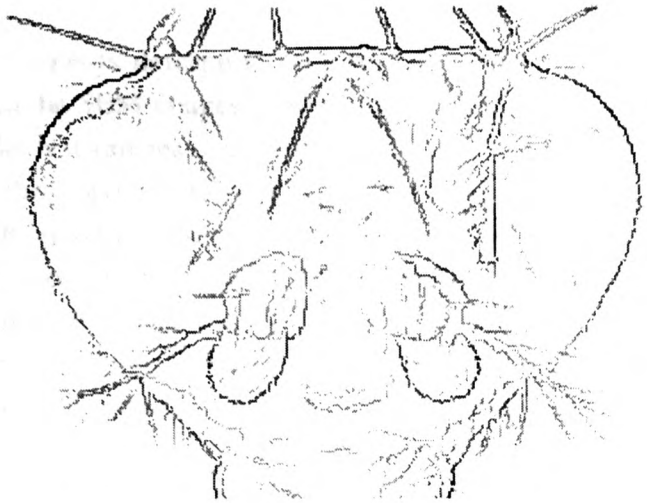


Fig 4.6. Final contour map for the image in figure 1.1(a) after thresholding and merging the contour responses at each scale.

The second image analysed was the synthetic image shown in figure 4.7(a). This contains both contour and textural features of varying spatial frequency. The resulting contour map is shown in figure 4.7(b). The results for this image are promising in that the textured regions of different spatial frequency can be eliminated while retaining the important contour features. Discontinuities arise within certain edges as a result of the interaction between maxima in close proximity. Features on the perimeter of the textured region are also identified as contours. This results from the decay in interaction around the border of textured regions which neighbour continuous regions.

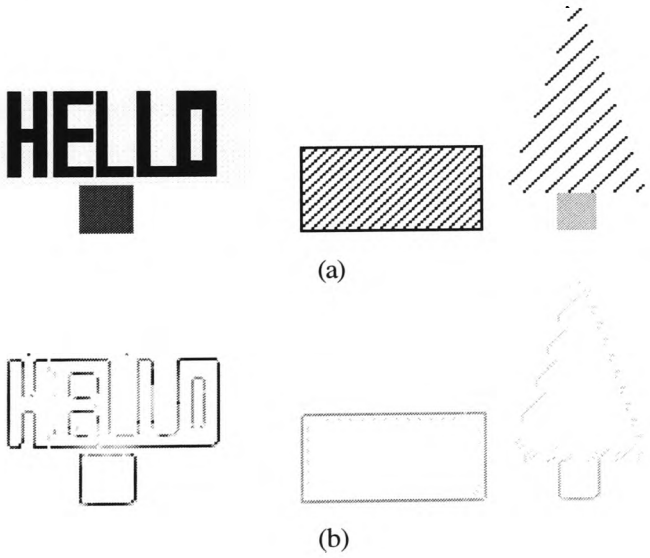
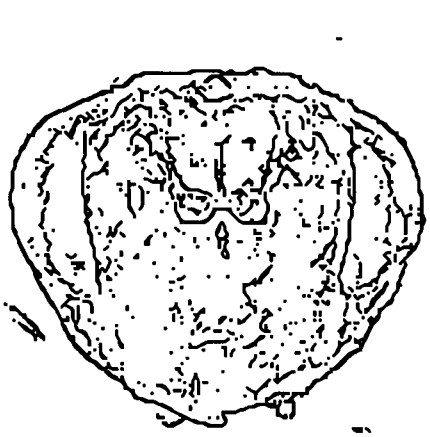


Fig 4.7. (a) Synthetic image containing contour and textural features of varying spatial frequency. (b) Resulting contour map with textural clutter filtered out while important contour information has been retained.

The next set of experiments determine how well the above contour / texture delineation algorithm performs on the SEM images in appendix 1. The above experimental parameters are also used here and the test images used from appendix 1 are {e100, e200, e300, e400, e402, e500}. As above, the wavelet transform is created and the contour response for each maximum in the multi-scale representation is then calculated. This is used to form the final contour map for each experimental image. The original image data is shown in appendix 1 and the associated wavelet transforms can be found in appendix 4. Figure 4.8 shows the final contour map for each test image. These results have been transformed into binary images to highlight those points that are retained in the final contour maps. The problem here is that significant textural clutter remains. The main reason for this can be found in the spatial properties of the texture, where textural features reside at larger scales of analysis within the wasp head images. This can clearly be seen from the wavelet transform data shown in appendix 4.

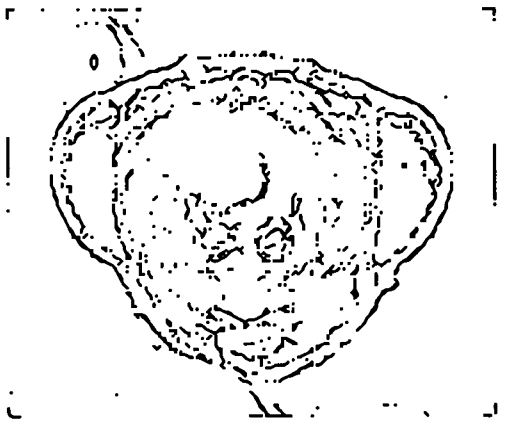
The anterior view of the Fruitfly head shown in figure 1.1(a) contains textural features which reside at smaller scales of analysis. As such, the resulting contour map in figure 4.6 contains significantly less texture than the contour maps for the wasp head images shown in figure 4.8. Using the ROI curve defined in (22) and shown in figure 4.3, interaction between maxima and their neighbours at the largest scale of analysis will be calculated within a 1x1 pixel region. This results in no interaction around each maximum and hence all edge events at the largest scale of analysis will have a high contour response, resulting in significant textural clutter surviving in the final contour maps. One solution to this problem would be to increase the scale range so the ROI curve defined in (22) will be sufficiently large at the scale $\sigma=2^3$. However this is computationally expensive, both in terms of processor time and storage as we would need to calculate the wavelet transform over a greater scale range. Alternatively, the ROI curve could be expanded to allow the interaction between maxima at larger scales. This is equivalent to using the original curve defined in (22) over a larger scale range, only the expense of calculating and storing the augmented wavelet transform is eliminated. Figure 4.10 shows how the ROI curve can be modified to force maxima to interact at larger scales of analysis and shows the relationship between ROI curves defined over a larger scale range. The modified ROI curve shown in figure 4.10 has been used to create contour maps using the test images listed above. The same experimental conditions remain and only the ROI curve has been altered. The resulting contour maps, again represented as binary images, are shown in figure 4.9.



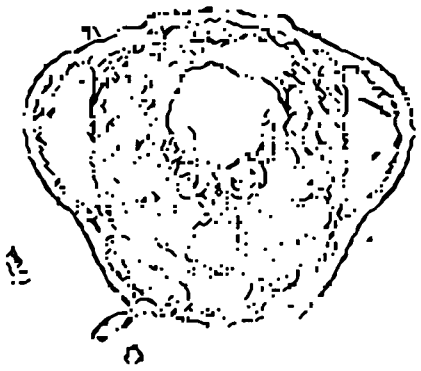
Experiment : e100

Experiment : e200

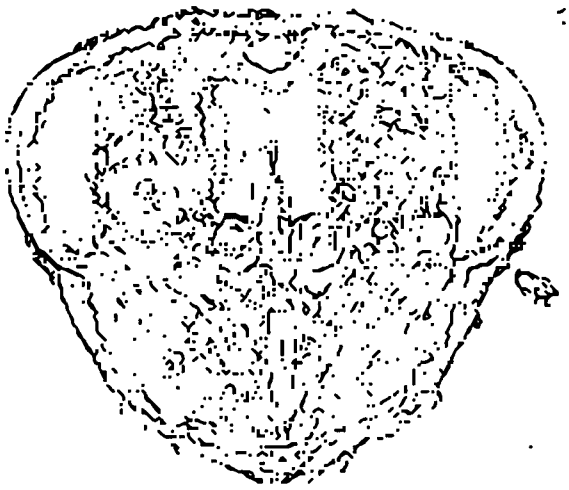
Experiment : e300



Experiment : e400

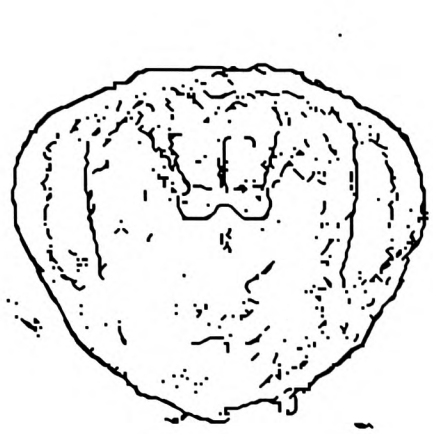


Experiment : e402



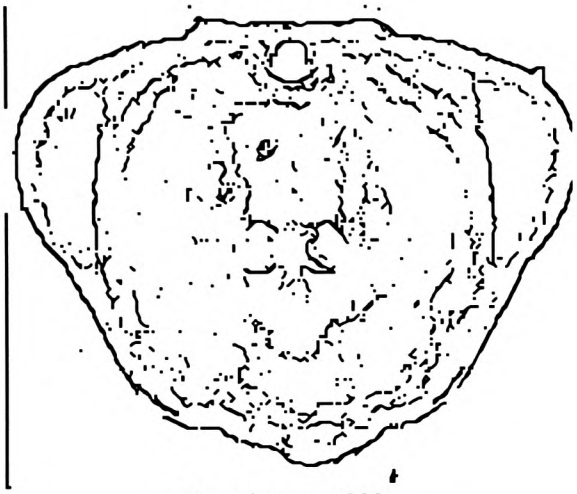
Experiment : e500

Fig 4.8. Contour maps created for the test images in appendix 1. Significant textural detail remains.

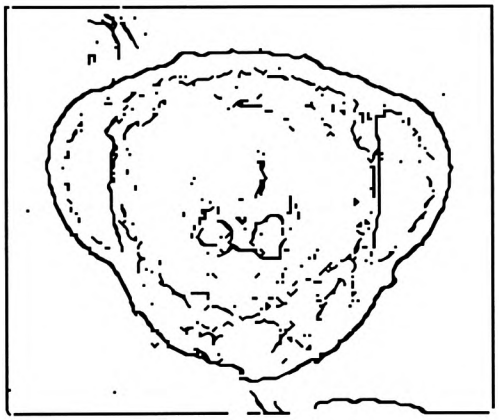


Experiment : e100

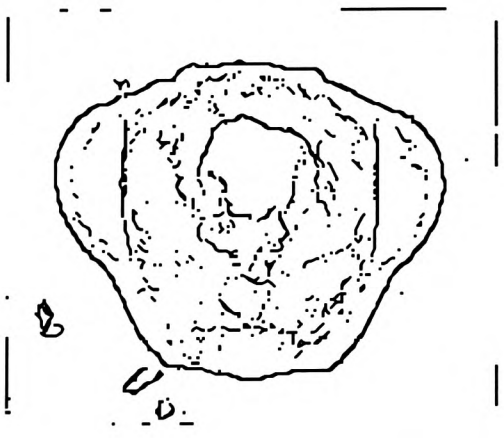
Experiment : e200



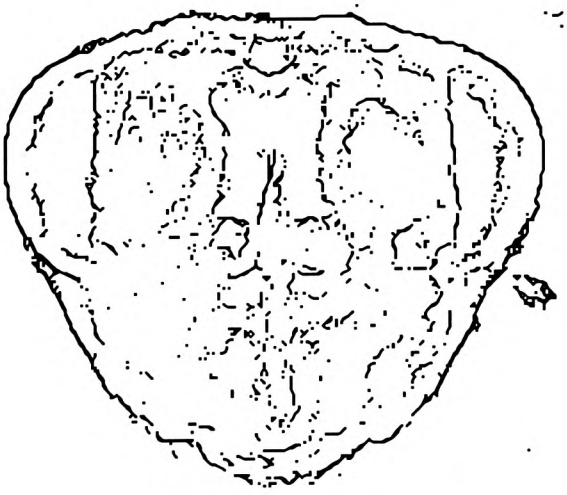
Experiment : e300



Experiment : e400



Experiment : e402



Experiment : e500

Fig 4.9. Contour maps created using a modified ROI curve. Textural clutter is significantly reduced.

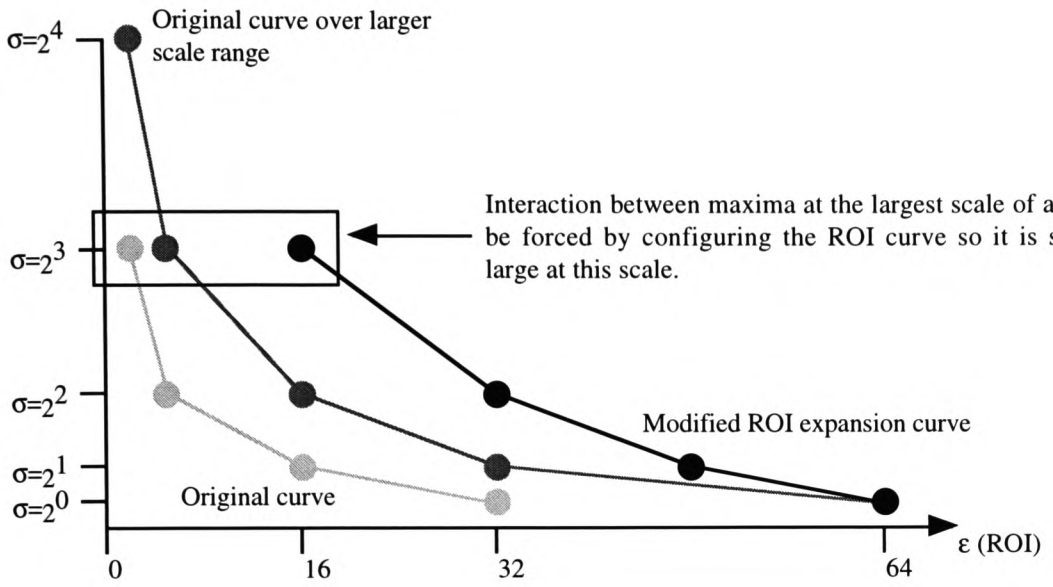


Fig 4.10. ROI expansion curves can be configured within a confined scale range so maxima at larger scales of analysis can interact with each other. The modified ROI curve shown here is defined as ε=(16,32,48,64)

These results show a significant reduction in the presence of textural clutter. Clearly the size of the region of interest over the larger scales of analysis is critical in eliminating unwanted textural clutter from the wasp head images. These results also show that a suitable expansion of the ROI in which maxima interaction is calculated can filter non-salient textural clutter and reduce the complexity of the resulting contour map. However, one problem that remains is the erosion of the contours that define the circle landmark features within each wasp head. This is caused by the fact that some instances of the circle landmarks are defined by two concentric circles whose edges lie in close proximity. This gives rise to the above problem where two maxima close to each other produce a large interaction value. This only occurs at certain segments of the circle boundaries and thus results in fragmented circles in the final contour map. It is thus important to configure the ROI curve so this erosion is minimised. Appendix 5 shows the final contour maps produced for each of the wasp head images in appendix 1 using the ROI curve ε=(32,64,128,128). The faster expansion of this curve further eliminates textural clutter from the resulting contour maps while minimising the erosion of salient edge features.

4.5. Conclusion

This chapter has studied the problem of deriving a suitable representation from the Mallat wavelet transform for subsequent landmark localisation and species classification. No one scale of analysis is suitable because salient contour features erode at larger scales while significant textural clutter exists at smaller scales. Thus the aim of this work is to reduce the amount of data created in the wavelet transform by recombining the multiscale representation into a single contour map which maximises the presence of contour features and minimises the presence of textural clutter. The context in which the image is analysed, the wavelet transform modulus maxima and the local activity around each maximum are deemed important features. These have been used to create contour maps for the wasp head images in appendix 1. The results show that significant textural clutter can be filtered out, provided the ROI curve used to calculate maxima interaction is sufficiently large in the upper scale range where irrelevant textural clutter resides in the wasp head images. For other images this is not so critical because textural clutter resides at smaller scales of analysis. The resulting contour maps can now be used for subsequent processing to locate the taxonomic landmark points shown in appendix 2. This work represents a significant part of the contribution to knowledge and a novel way of processing textured images using edge detection methods. First it allows a single contour map to be derived from highly textured images, overcoming the problems of poor edge localisation and the introduction of correlated noise inherent in texture analysis methods. Second the localisation of edges in the final contour map is excellent since all edges can be traced back to their true location in the wavelet transform. This overcomes the poor localisation results found in Ganesan *et al.'s* work [38] and is not restricted to Davis' simple structural textures [25]. The resulting contour map can be used for building intermediate and high level shape models for subsequent object and landmark localisation.

Chapter 5

Shape Representation

5.1. Introduction

The previous chapter describes how the Mallat wavelet transform can be used to derive a single contour map where salient edge features are maximised while irrelevant textural clutter is minimised. Given this intermediate representation, the next stage is to extract meaningful information about the objects, or sub-structures on the wasp head which contain important landmark features. As described in chapter 1 and illustrated in appendix 3, the sub-structures which comprise the wasp head either correspond to whole landmark features or contain landmark points on their respective boundary. First, post-processing of the "texture filtered" contour map is described, where contour segments are extended to overcome the problem of fragmented edges inherent in the Mallat wavelet transform described in chapter 3 and the texture filtering process described in chapter 4. Unfortunately, unwanted textural clutter can be reintroduced into the contour map, so a novel technique is presented which uses the statistical properties of contour segment lengths to identify reintroduced textural segments and eliminate them. This technique also represents an important contribution to knowledge, augmenting the texture filtering technique described in chapter 4. The hierarchical representations used to model both the wasp head sub-structures and the wasp head as a whole are then discussed. Given the limitations of deformable contour models in the presence of local image clutter, this chapter focuses on dominant curvature point representations as the basis for these hierarchical models. Finally, methods for encoding the deformation between instances of each wasp head and their respective sub-structures are discussed. A novel algorithm is presented which automatically finds correspondences between dominant curve points that define deformed instances of an object. This overcomes the problems of having to specify exact point correspondences as in techniques such as the ASM (Active Shape Model). The resulting statistical model which is described encodes the deformation for the wasp head and its component sub-structures.

5.2. Contour Model Extraction

Calculation of the Mallat wavelet transform results in discontinuous contour features at larger scales of analysis. This results from the splitting and merging of edge points across scale space and the fact that a 1:1 correspondence between maxima at adjacent scales does not necessarily exist. These discontinuities propagate into the final contour map as shown in the results of chapter 4 and appendix 5. This problem can be overcome by using the more complete edge information in the wavelet transform. Since the smallest scale of analysis contains all of the edge data in the original image, this will be used to augment the incomplete edges in the contour map. However, two important problems must be overcome. Firstly, extending contour segments using a cluttered edge map is likely to produce poor results as

erroneous contour extensions may result [64]. This problem is relevant here because the smallest scale of analysis contains all of the unwanted textural features. Secondly, since residual "texture" is retained in the contour map, surviving textural elements may be extended, reintroducing unwanted textural clutter. To overcome these problems, a constrained contour extension algorithm is described which uses the direction of the contour and maxima intensity to guide the search for neighbouring edge points in the wavelet transform. The statistical properties of the augmented contour segments are then considered to filter out those segments which correspond to reintroduced textural clutter.

5.2.1. Contour Extension Methodology

Given the problems of augmenting contour segments using a cluttered edge map, it is important that the contour extension techniques adopted are suitably constrained. Here contours are simplified and considered as 1D structures which reside in the 2D image plane. As a result no branch points can exist so no two edges can converge in the image plane. Given this model, each point can have a maximum of two neighbouring edge points. The first stage of the contour extension process therefore ensures that the contour map representation adheres to this model. To achieve this, the morphological erosion operator [73] is applied to the contour map, eliminating points where more than two neighbouring edge points exist. Figure 5.1 illustrates how the erosion of an example edge is necessary to enforce the above 1D contour model.

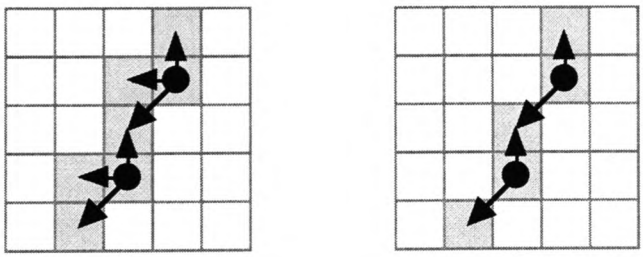


Fig 5.1. (a) The contour map may contain points where more than two neighbouring edge points exist. (b) Erosion of the contour map eliminates erroneous points so all remaining edge points have at most two neighbours.

The second stage of the process creates links between edge points in the contour map. Each point can be classed as (i) an "isolated" point, where no neighbouring edge points exist, (ii) a "terminal" point which resides at the end of a contour segment and has only one neighbouring edge point (although this status is actually only allocated to points at the end of contour segments which cannot be extended further in the algorithm described below) and (iii) a "complete" edge point, where both neighbouring edge points exist. For each edge point a link

vector L is stored where L=(n1,n2). Elements n1 and n2 represent the offsets to the neighbouring edge points in the image plane, where n1=(x1,y1) and n2=(x2,y2). If an edge point is classed as "terminal" or "isolated" then either both or one of the links respectively will be set to NULL, ie. n_i=(0,0). Figure 5.2 illustrates how these links are created for each of the above edge point classes.

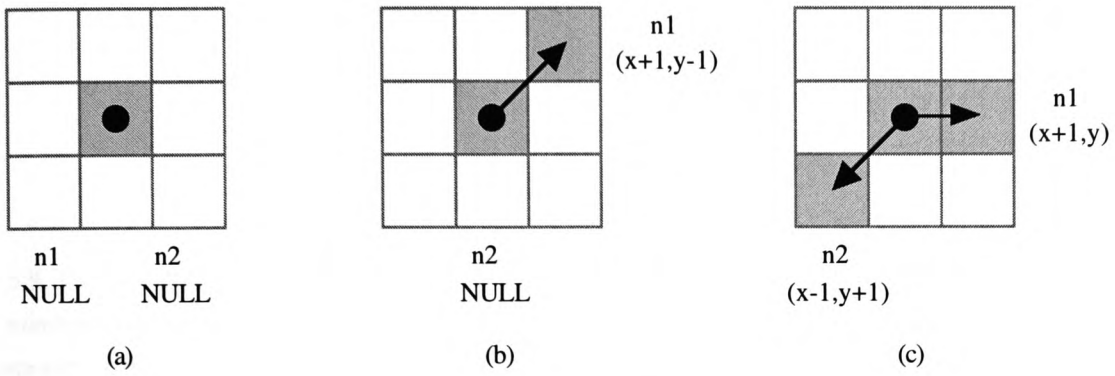


Fig 5.2. (a) An "isolated" point has no neighbours so both links are NULL. (b) A "terminal" point has only one neighbour, so one link points to this while the other is set to NULL. (c) A "complete" point has two neighbours so both links are setup to point to these.

Once the contour map has been eroded so it complies with the 1D contour model described above and once the links between neighbouring edge points have been created, the final stage of the process performs the actual contour extension. This works by first looking for points in the contour map which have only one neighbouring edge point. Local edge information in the wavelet transform is then analysed to determine if the associated contour segment can be augmented. For each point in the contour map that satisfied the above one neighbour constraint, the number of neighbouring edge points in the wavelet transform are counted. The maximum in the wavelet transform which corresponds to the current point's only neighbour in the contour map is ignored. This is illustrated in figure 5.3.

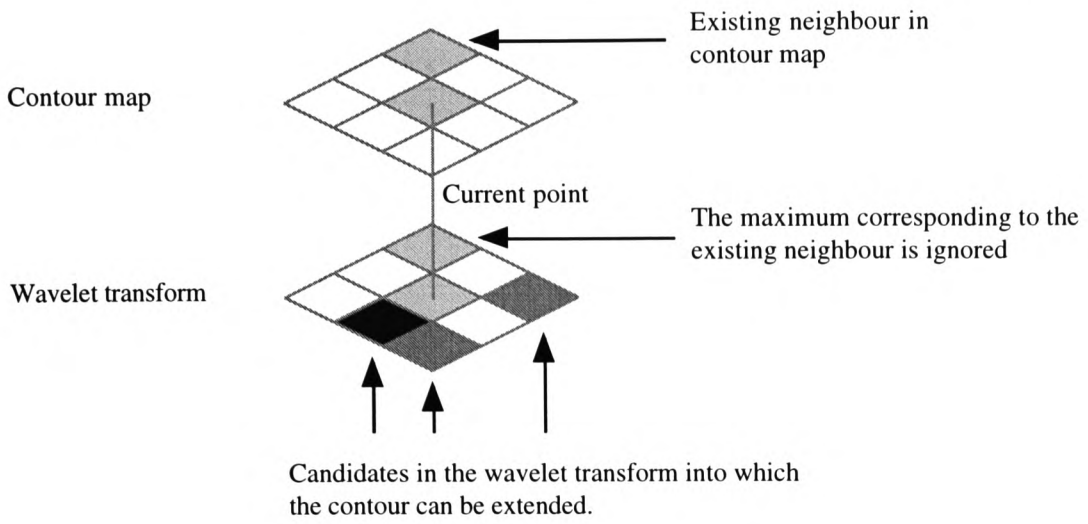


Fig 5.3. For each point in the contour map which has only one neighbour (the centre point on the above grid), the number of neighbours in the surrounding wavelet transform are counted. Here only 3 neighbours exist because the maximum corresponding to the existing neighbour in the contour map is ignored.

If no neighbouring maxima exist in the wavelet transform then the current point is marked as a "terminal" point since this represents the end of the contour segment in both the contour map and wavelet transform. If only one neighbouring maximum exists in the wavelet transform it is subsumed by the respective contour segment and included in the contour map representation. Links are created between the new and current points and their status is adjusted accordingly, where the current point now has two neighbours while the new point has only one. Figure 5.4 illustrates this process. Hysteresis thresholding is then applied to the new edge point to determine if a noise-defined edge is being augmented [13]. If so, the new point is marked as a "terminal point" and cannot be extended further.

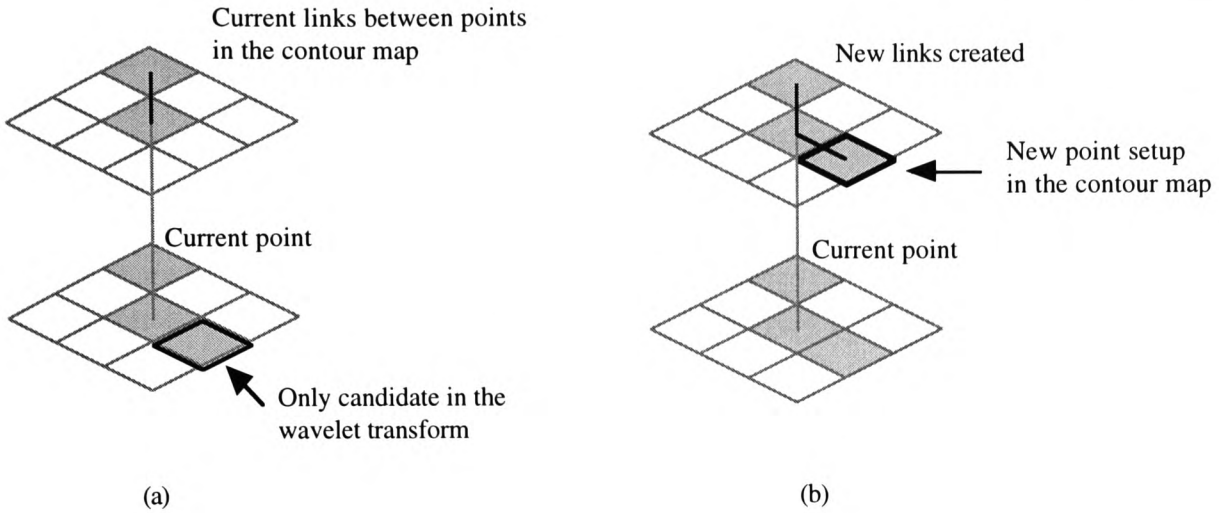


Fig 5.4. (a) Only one neighbouring maximum exists in the wavelet transform. (b) This is included in the contour map and new links with the current point (in the centre of the grid) are created.

If more than one possible neighbour exists in the wavelet transform, the local direction of the contour and the maxima intensity is then used to determine the most likely candidate. Figure 5.5 shows how an edge point can be selected from the wavelet transform where more than one neighbouring maximum exists. This process continues until a stable state is reached where all of the edge points in the contour map which have only one neighbour are classed as "terminal" points and cannot be extended further.

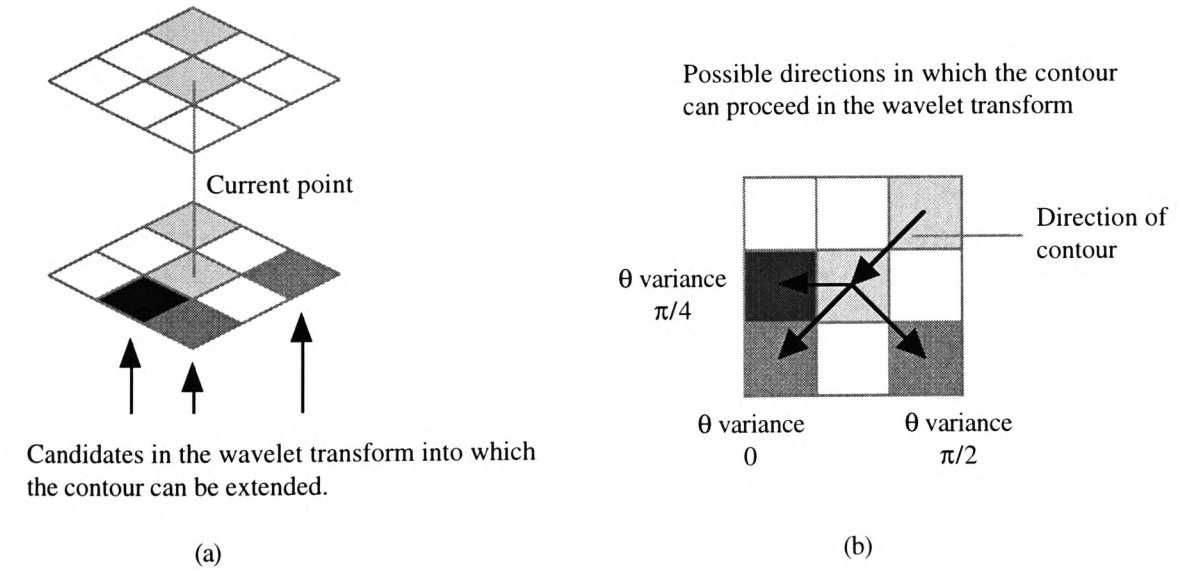


Fig 5.5. (a) More than one neighbouring maximum exists in the wavelet transform. (b) Select the best candidate based on (i) the orientation with respect to the local direction of the contour and (ii) maxima intensity. Maxima with the least orientation and intensity variance are selected.

The above algorithm has been applied to the contour maps e100,e200,e400 and e500 from appendix 5. The resulting augmented contour maps are shown in figure 5.6. It is evident from these results that important edges which define the main sub-structures on the wasp head do not diverge given the clutter in the wavelet transform and discontinuities have been eliminated. However, significant textural clutter is reintroduced, as was previously noted. The algorithm described above therefore overcomes the problems of using a cluttered edge map to augment discontinuous contour segments but it is not capable of distinguishing between contour and textural features.

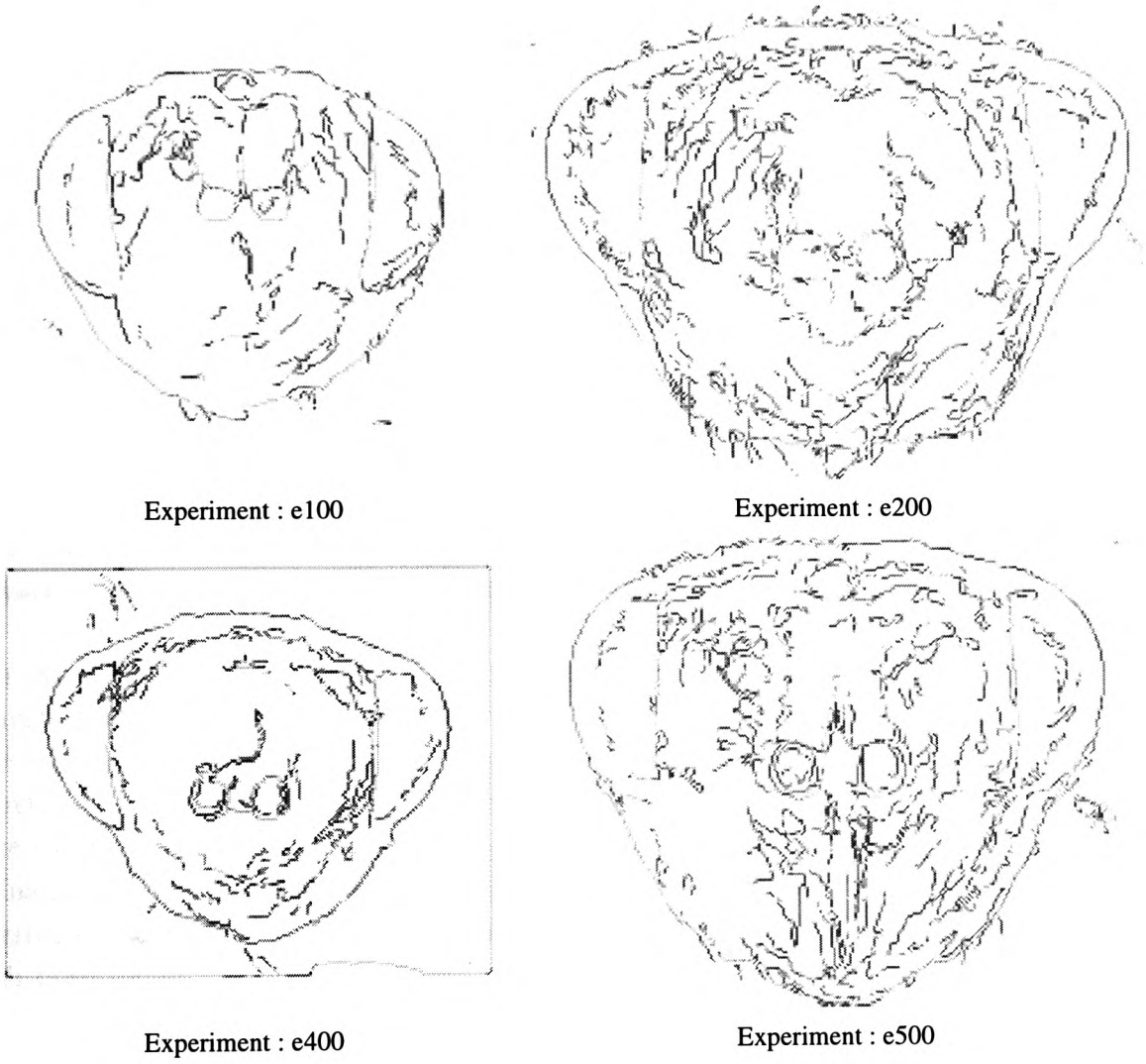


Fig 5.6. Contour extension results. Textural features are reintroduced.

5.2.2. Statistical Properties of Edge Elements

Augmenting contour segments which correspond to large scale textural features results in the reintroduction of textural clutter as shown above. Hysteresis thresholding is applied to prevent noisy edges from being reintroduced. However this does not prevent textural features from being augmented. This can be seen in the edge strength profiles shown in figure 5.7. Here the strength of an actual edge and a reintroduced textural segment do not fluctuate significantly, nor does the textural edge fall below the hysteresis threshold.

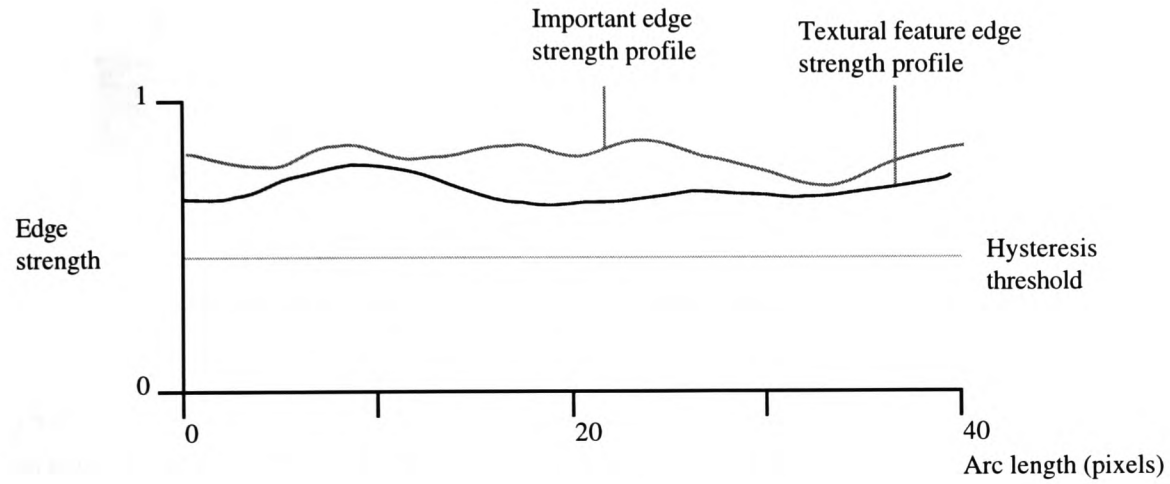


Fig 5.7. The edge strength profiles of an actual contour and the contour that defines a textural feature shows that hysteresis thresholding is not suitable for preventing the reintroduction of textural clutter.

From the results shown above, it is apparent that the contour segments reintroduced to eliminate discontinuities in important edge segments are short while the contour segments corresponding to textural features are much longer. This is because the intermediate contour map retains a significant amount of contour information used to define the structures of interest on the wasp head, as shown in appendix 5, and only small contour extensions are required to eliminate the discontinuities introduced in the wavelet transform. Therefore the statistical properties of the reintroduced contour segment lengths can be used to filter out those segments which correspond to textural features. An appropriate threshold can be selected and reintroduced segments whose lengths are greater than this threshold can be eliminated. This approach is used to prevent the reintroduction of textural clutter shown in the above results.

In order to calculate the lengths of each contour segment, we must partition the contour map so each individual segment can be extracted. Just as image segmentation assigns a label to each region, so a label will be assigned to each contour segment. This is done by initially assigning each edge point with a unique label, or integer value. Each edge point then communicates its

label to neighbouring points via the link structure L=(n1,n2) described above. If a neighbouring point has a smaller value label then it takes on the larger value label of the current edge point. This process is applied to all of the edge points in the contour map and is repeated until a stable state is reached, where the labels do not change from one iteration to the next. In order to discriminate between original contour segments and those created with the above contour extension algorithm, communication of the edge point labels between new / reintroduced and original contour segments is prevented. This process is illustrated in figure 5.8.

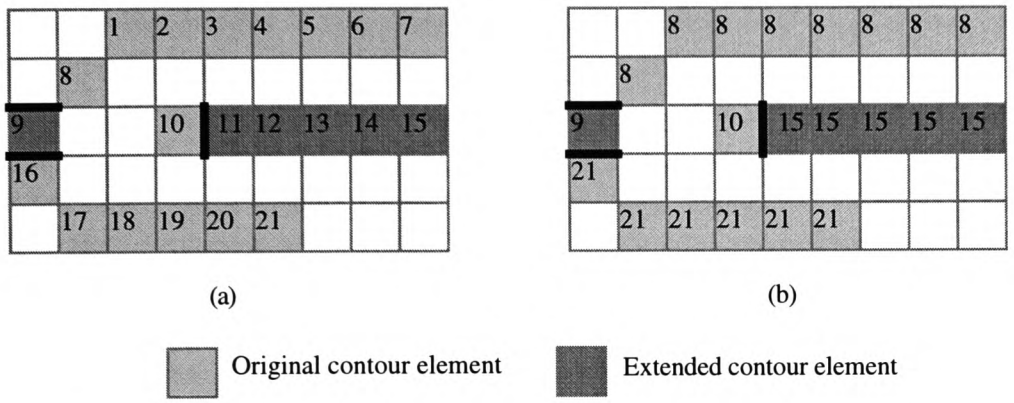


Fig 5.8. (a) Initially all edge points are assigned a unique label, or value. (b) Each edge point communicates its label to its neighbours (communication between old and new/extended contour segments is prevented). Smaller value labels are subsumed by larger value labels. This results in each contour segment having a unique label.

Once the contour map has been labelled, a table is created storing the lengths of each new contour segment. Each segment is indexed by its associated label and its length is simply the number of edge points in the contour map that have the same label. Table 5.1 lists the labels and segment lengths for the example illustrated in figure 5.8 (b).

Table 5.1.
Reintroduced Contour Segment Lengths

Contour Segment Label	Length
9	1
15	5

From this data, we can calculate an appropriate threshold. Let the mean contour length l be defined as

$$l = \frac{1}{n} \sum_{i=1}^{i \leq n} \text{Clength}_i \tag{31}$$

where n equals the number of new contour segments added to the contour map using the above extension algorithm and Clength_i represents the length of contour segment i. In the above example, the number of new segments, n, equals 2. The threshold is set equal to the mean segment length. Any reintroduced (extended) contour segment whose length is greater than this threshold is then removed from the contour map. Figure 5.9(a) shows the resulting contour map for the example given in figure 5.8. where the reintroduced segment whose length is greater than the mean length of 3 has been filtered out. However, the surviving contour is comprised of three segments (8,9,21). This is because of the constrained labelling operation described above, where old and new / reintroduced contour segments could not communicate their labels to each other. This problem is easily overcome by applying the same labelling algorithm to the new contour map but now no labelling restriction exists. The resulting contour map is shown in figure 5.9(b). In the same way new contour segments could be extracted using the label data, so complete contours can now be extracted for subsequent use in building high level models of the wasp head and its sub-structures. Table 5.2 lists the contour structures contained in the final contour map shown in figure 5.9(b).

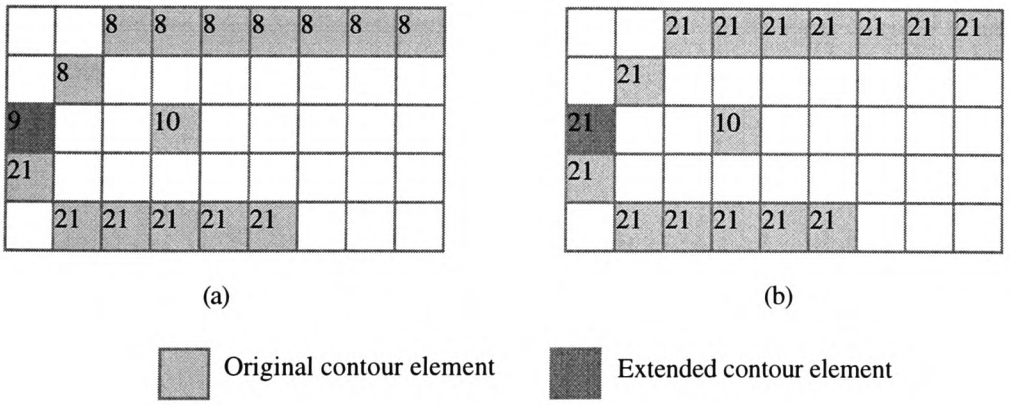


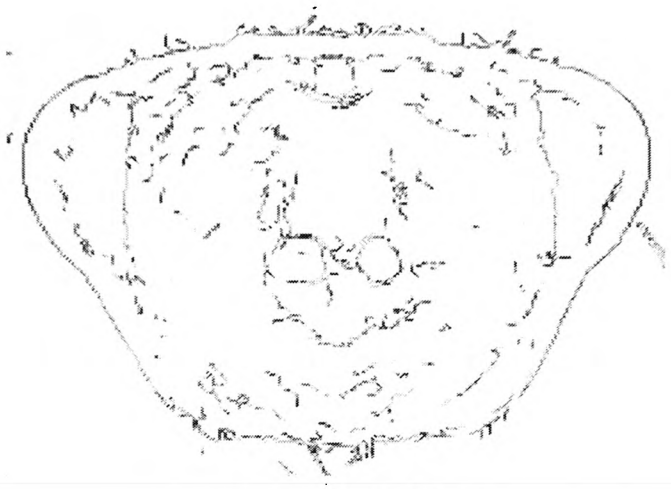
Fig 5.9. (a) Filtering reintroduced contour segments whose length is greater than the mean segment length eliminates possible textural clutter. Here segment (15) has been eliminated. (b) Relabelling without the above restriction between old and new /extended contour segments allows whole contours to be formed.

Table 5.2. Final Contour Structures	
Contour Segment Label	Length
21	15
10	1

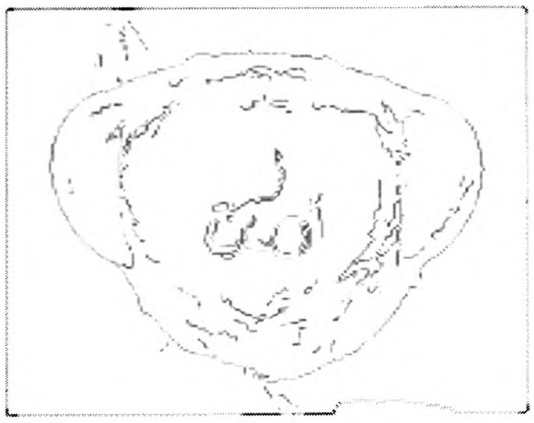
To test the effectiveness of this approach on the wasp head image data, the above labelling and filtering algorithm is applied to the experimental images e100, e200, e400 and e500 used above. Figure 5.10 shows the results of using the mean segment length threshold (l), figure 5.11 shows the results of using a threshold below the mean (l / 2) while figure 5.12 shows the results of using a threshold above the mean length ($l + l$ / 2).

The results show that when the threshold is above the mean segment length reintroduced textural clutter is retained. When the threshold is below the mean segment length important edge information can be filtered out. However, when the threshold is equal to the mean segment length a good balance is achieved. Important contour information is retained while the reintroduction of textural clutter is significantly reduced. Therefore, filtering contour segments with the mean contour segment length threshold produces suitable results for deriving high level contour models. Some discontinuities remain in the final contour maps, but this is unavoidable given the above contour extension process must be constrained to avoid the divergence of contour segments as a result of the clutter in the wavelet transform. There are some cases where the mean length can be affected by image data, as in the case of image e105 shown in appendix 5. Here the border around the head resulting from background image clutter causes the mean segment length to be much higher. As such, more texture is retained in the final contour map.

Experiment : e100



Experiment : e200

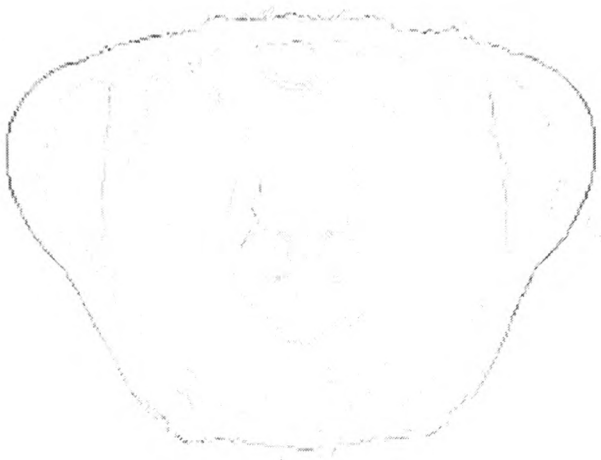


Experiment : e400

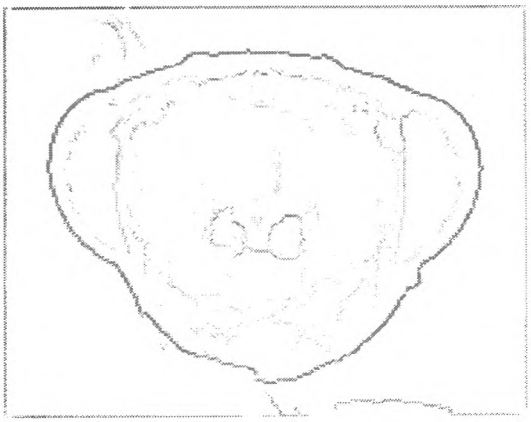
Experiment : e500

Fig 5.10. Filtering contour segments where the threshold equals the mean segment length.

Experiment : e100



Experiment : e200



Experiment : e400

Experiment : e500

Fig 5.11. Filtering contour segments where the threshold is less than the mean segment length.

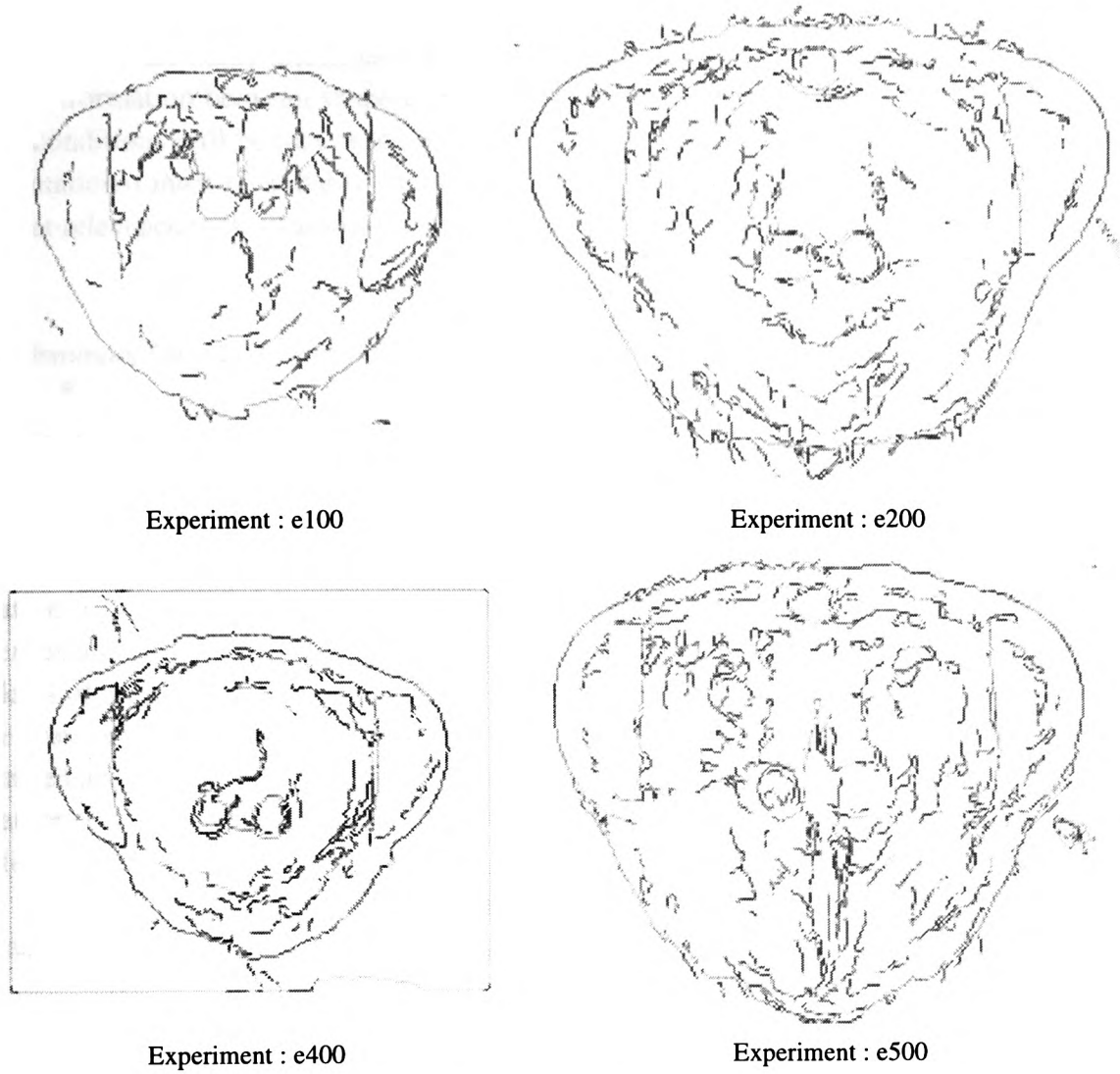


Fig 5.12. Filtering contour segments where the threshold is greater than the mean segment length.

5.3. Wasp head and Sub-structure Representation

This section discusses the representational issues relevant to modelling the wasp head and its sub-structures for subsequent landmark localisation. As shown in appendix 3, the wasp head is decomposed into component objects which either correspond to whole landmark features (circles C1 and C2) or contain landmark points on their respective boundaries (the eye object E1, the top of head object T1 and the bottom of the head object B1). Ideally each object would be represented as an open or closed 1D curve function which stores the curvature at each point along its length. However, the contour map described above does not lend itself to this idealised representation because of the discontinuities that remain between contour segments and the presence of residual textural clutter. Hence the formation of closed curves from

fragmented contour segments is made difficult if not impossible. Aside from 1D curve functions, there are many techniques described throughout the literature for modelling shapes and the deformation between shape instances. It is important to note that we cannot assume certain conditions will be met in the input data and as such post-processing algorithms (and representations) must be suitably robust to handle variable and noisy input data. This is of particular relevance to the wasp head data being used here.

5.3.1. Dominant Curvature Point Representation

Since the presence of discontinuities and moderate textural clutter in the contour map makes the use of contour segment based representations infeasible, it is considered that dominant curvature point representations [104] may lend themselves to the modelling of the wasp head sub-structures. Since most of the information about an objects shape is stored at points of high curvature [6], a complete curve segment representation is inherently redundant. There are numerous algorithms that find the points of high curvature on 2D planar curves. However, problems of noise in the contour can given rise to erroneous "corner" points. Curvature scale space can overcome this by smoothing the curve function at each scale. Noise can be eliminated and the general curvature properties of the object can be extracted. The properties of curvature scale space are the same as those for the multiscale representation described in chapter 3. Based on the gaussian smoothing function, points of high curvature can be derived from the modulus maxima in the curve function itself or from zero crossings in the first derivative of the curve function.

The maxima in the Mallat wavelet transform had advantages over zero crossings in that important features could be distinguished from irrelevant ones. For curvature scale space however, the use of maxima introduces some important problems. First, the value of the maxima determines the amount of local curvature about that point. Unfortunately, the curve function deforms at larger scales of analysis, just as edges move, split and merge in the Mallat wavelet transform. Figure 5.13 shows an example curve at two different scales where the curvatures at corresponding points do not match. The maxima values across curvature scale space thus do not accurately represent the true curvature at the respective edge point.

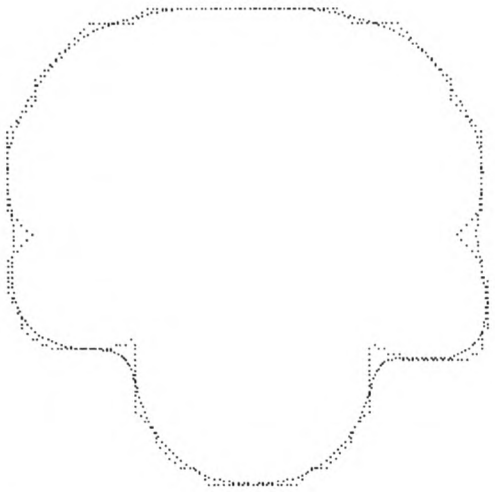


Fig 5.13. Contours and the resulting curve function deform over curvature scale space.

Secondly, local maxima or minima in the curvature function indicate whether the curve is convex or concave at that point respectively. Figure 5.14 shows the curvature functions for the above shape where each function is calculated from the boundary point order proceeding in clockwise and anti-clockwise directions. Here the concavity / convexity of the dominant curve points are inverted and are thus dependent upon the point ordering of the 2D planar curve from which the curvature function is calculated. Given the fragmented contour segments in the contour map, we cannot guarantee the point ordering on each segment will always be consistent.

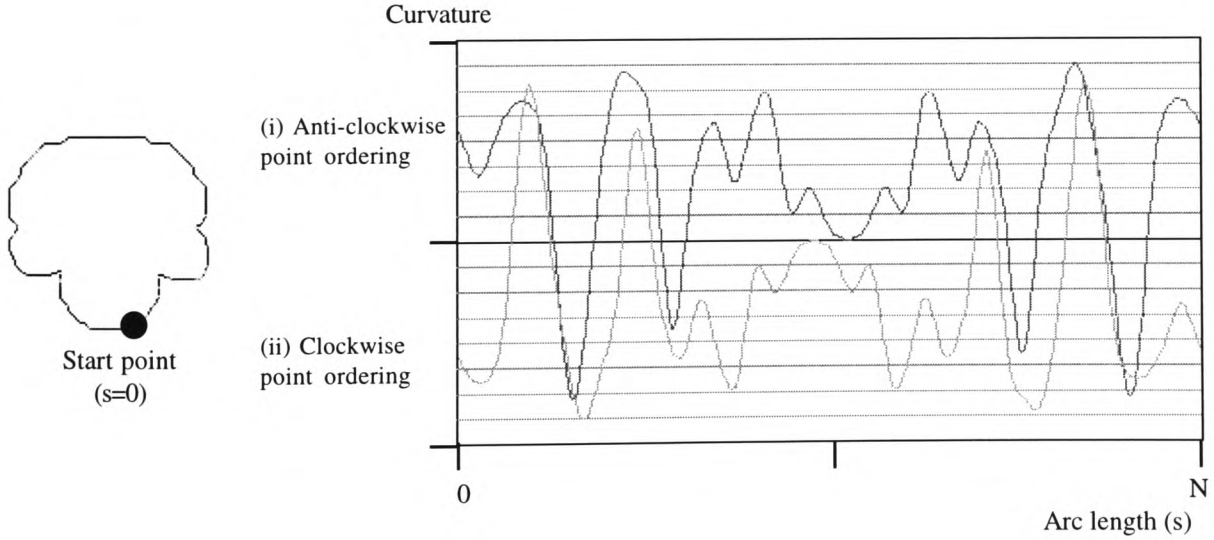


Fig 5.14. Curve functions become inverted when the point order on which they are calculated is reversed.

The final problem of using maxima to represent dominant curve points in curvature scale space is related to the part-whole criteria. This determines the relationship between an isolated curve segment and its corresponding segment in a complete curve. For sufficiently small curve segments where the width of the gaussian smoothing function is larger than the length of the segment or the smoothing function goes beyond the boundary of the segment, maxima around the end points of the curve do not accurately represent the true curvature properties. This is because of missing curve information at the boundaries of the contour segment. Hence the periodicity of the curve segment must be set appropriately. If there is zero or 1N periodicity, where N is the length of the contour segment, significant contour deformation results over curvature scale space. This can be seen in figure 5.15(a) and figure 5.15(b). If 2N periodicity is used, where the contour segment is symmetrical at its boundary points, the contour segment does not deform significantly over the scale range and only changes shape as a result of the blurring caused by the gaussian smoothing function. The fidelity of the resulting curvature scale space is thus retained. This is illustrated in figure 5.15(c).

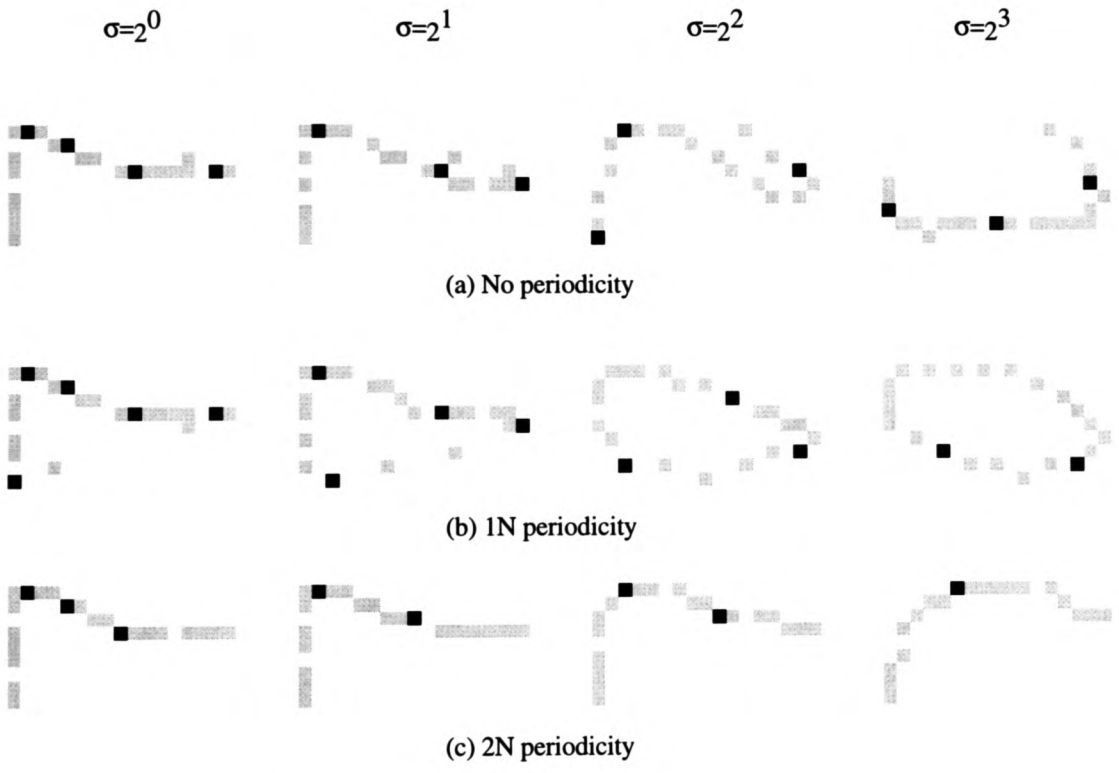


Fig 5.15. Deformation of partial contours over curvature scale space when there is (a) No periodicity, (b) 1N periodicity and (c) 2N periodicity. The shape deforms significantly for zero and 1N periodicity but only deforms slightly for 2N periodicity.

Unlike maxima, zero-crossings only retain positional information and as such are not prone to the above problems aside from positional displacement. Mokhtarian and Mackworth [89] show that curvature scale space representations that model points of high curvature using zero-crossings retain the invariance, uniqueness and stability properties of shape representation. Therefore the dominant curve point representation for modelling the wasp head sub-structures will be based on zero-crossings. Although each contour segment can be reduced to a set of dominant curve points, each segment extracted from the contour map remains isolated and forming a complete shape from these isolated segments remains a difficult problem. This can in fact be overcome easily by mapping the dominant curve points from the local contour space back into the image plane. The problem of representing the concavity / convexity of points is overcome by storing the direction perpendicular to the direction of the contour in the image plane at each dominant curve point. Structures in the image are no longer a collection of fragmented contour segments but now become a set of vertex points P_{xy} where

$$P_{xy} = (P_\theta, P_\sigma) \tag{32}$$

where P_θ and P_σ represent the orientation of the dominant curve point and the scale at which it resides within curvature scale space respectively. This is a more convenient shape representation because we can process sets of discrete points rather than sets of contour segments of arbitrary length. The redundancy inherent in the contour segment representation has also been eliminated. Figure 5.16 shows examples of the dominant curvature point representation for wasp head images from appendix 1.

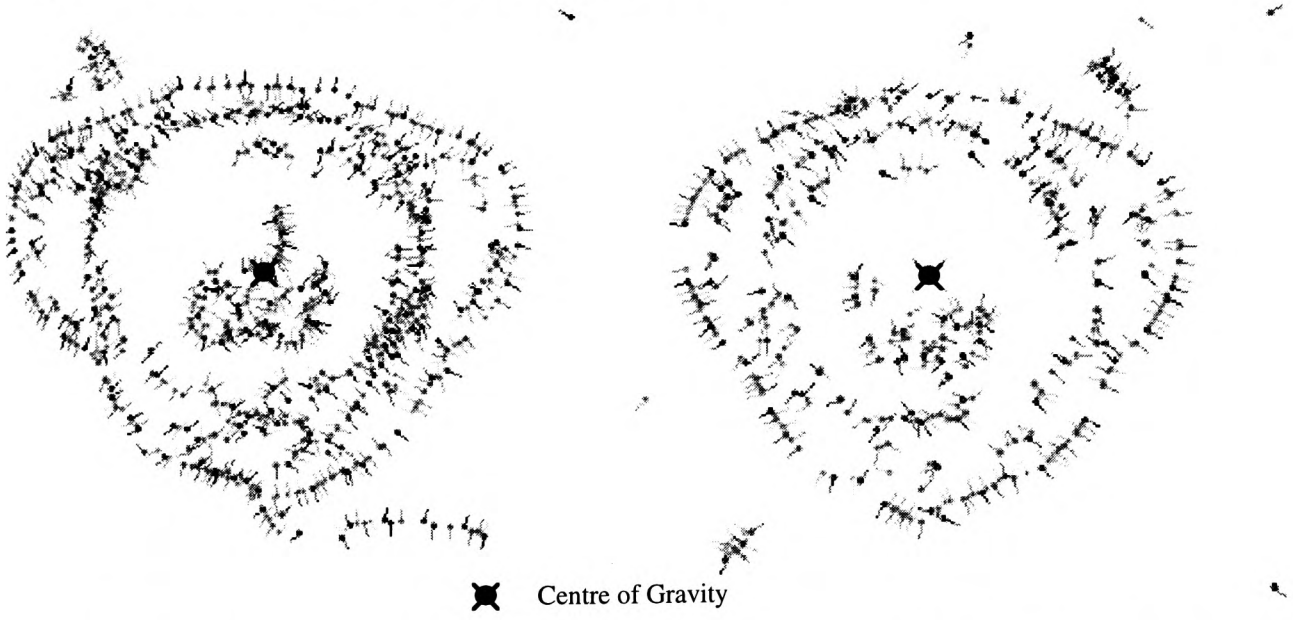
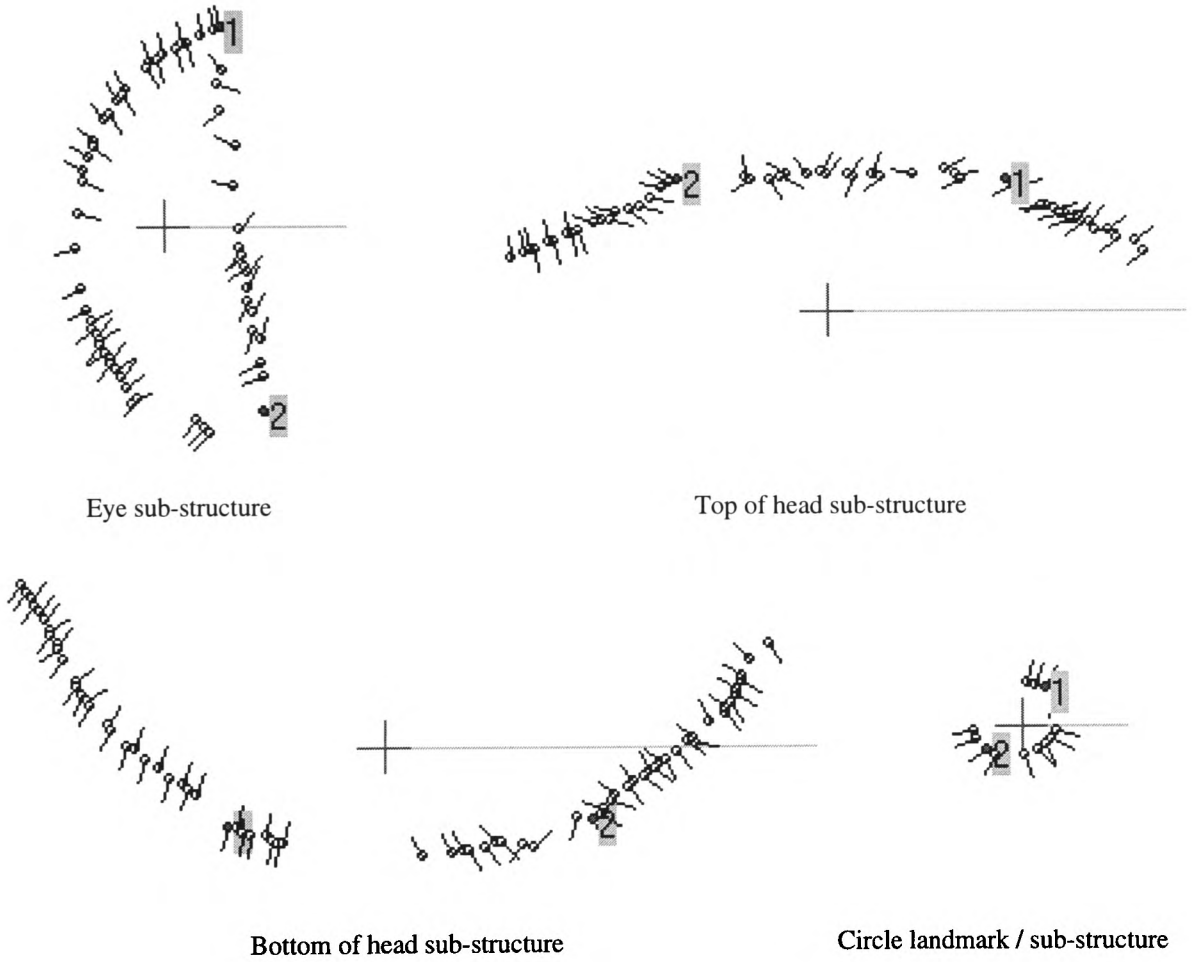


Fig 5.16. Example dominant curvature point representations for wasp heads e400 and e401 from appendix 1. Each dot represents a dominant curve point and each line represents the direction of curvature at that point.

Since each dominant curvature point remains isolated, we can model the wasp head sub-structures using a 2D polar representation. For each sub-structure the dominant curve points are stored relative to a common reference point, in this case the object centroid, or mean (x,y) position of its constituent points in the image plane. Figure 5.17 shows examples of each of the objects that comprise the wasp head model. Aside from the 2D representation, each shape is shown in its polar form along angular and radial axes. This representation is generic in that it is invariant to translation, rotation and scaling. If the shape was scaled, the representation is transposed on the radial axis. If the shape is rotated, then the representation is transposed along the angular axis. Normalising the polar representation eliminates these problems.



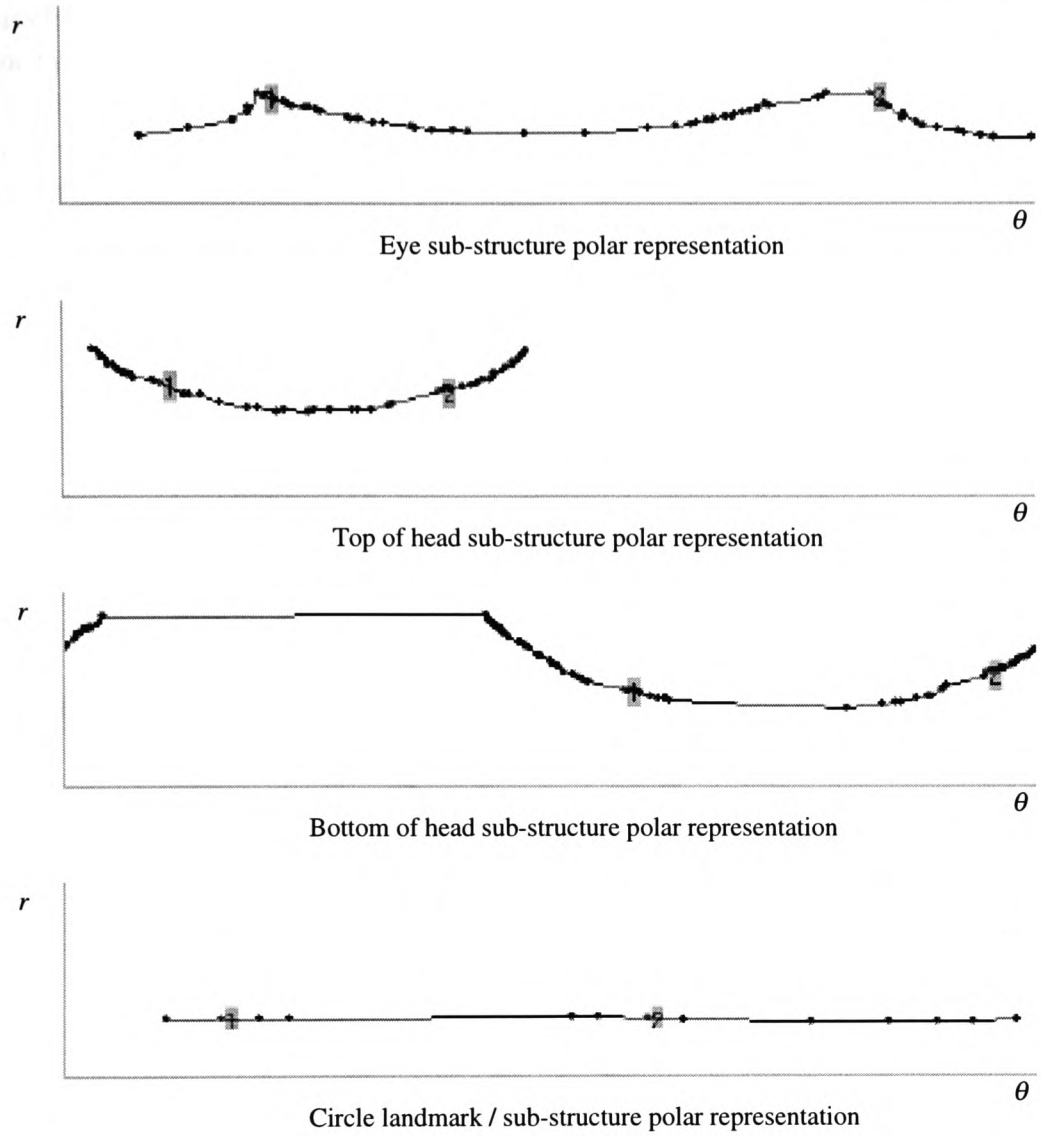
Eye sub-structure

Top of head sub-structure

Bottom of head sub-structure

Circle landmark / sub-structure

(a)

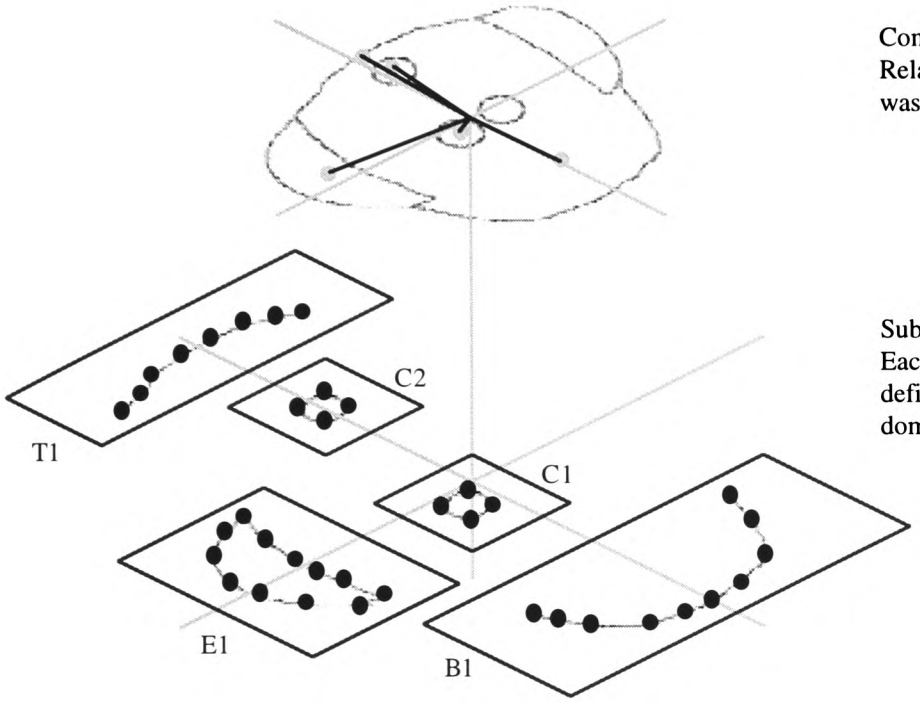


(b)

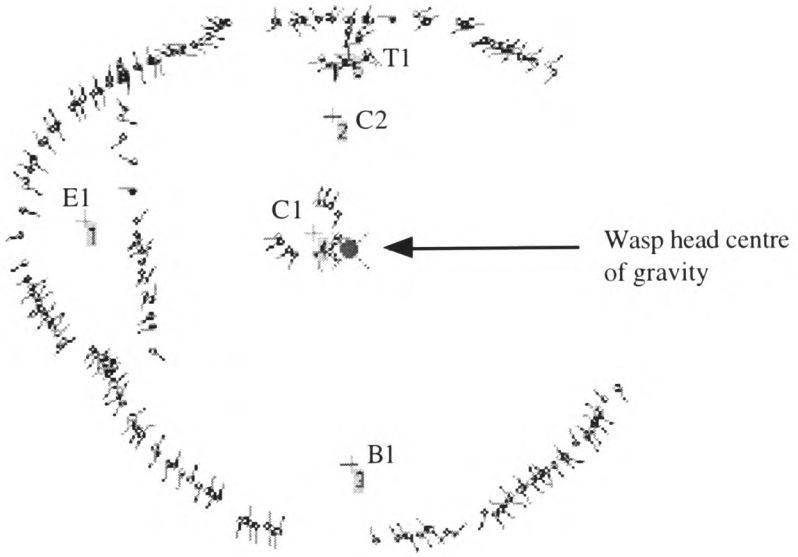
Fig 5.17. (a) Each sub-structure is represented as a set of dominant curvature points with respect to the object's centroid. The numbers in each diagram represent the taxonomic landmarks on each sub-structure. For the circle object these landmarks serve a functional purpose, as will be explained later, and have no taxonomic relevance whatsoever. (b) Polar representation of the above sub-structures, with the points ordered according to their orientation with respect to the object centroid.

Aside from the above sub-structures, high level knowledge about the wasp head as a whole is also important. This takes the form of modelling the spatial relationships between the sub-structures shown above which are relevant for landmark localisation. Other features on the wasp head extraneous to landmark localisation are not modelled here. Just as the dominant curve points are stored relative to a common reference point at the sub-structure level, so the

sub-structures at the wasp head, or compound object level are stored relative to a common reference point. This is the centre of gravity for the head, calculated as the mean (x,y) position of all of the dominant curve points derived from the original contour map. Figure 5.18(a) shows the wasp head representation hierarchy, illustrating the relationships between the compound and sub-structure levels while figure 5.18(b) shows an example of a stored wasp head model.



(a)



(b)

Fig 5.18. (a) Hierarchical representation showing the relationships between the wasp head sub-structures. (b) An example wasp head model showing the sub-structure centroids with respect to the head's centre of gravity.

5.4. Modelling Object Deformation

The above section described how the wasp head sub-structures can be modelled as a set of dominant curvature points within a polar coordinate representation. Although this model is generic and invariant to global object transformations, it is the natural deformation present in the wasp head image data that presents major problems and represents the main focus of this work. Figure 5.19 shows how two instances of the eye sub-structure can vary. Storing one instance of an object restricts the localisation methodology to fixed template matching as will be shown later. Since snake models are affected by image clutter, we need to model the deformation of objects using the dominant curve point representation described above.

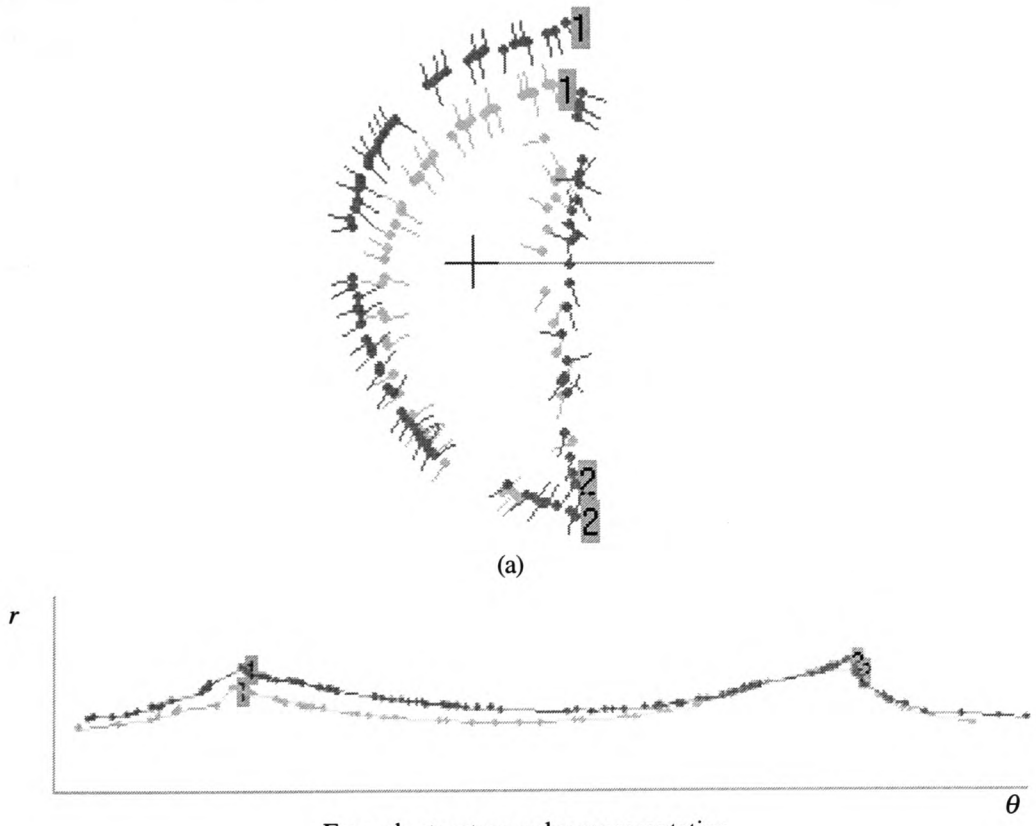


Fig 5.19. (a) Instances of the eye sub-structure can vary significantly. (b) This deformation results in the displacement of dominant curvature points in the polar representation.

5.4.1. Quantifying Distortion

For each object, a set of training instances are extracted from different images. Each instance I_i is defined as a set of dominant points ordered according to their orientation with respect to the object centroid.

$$I_i = < (\theta_1,r_1, P_\theta,P_\sigma), (\theta_2,r_2, P_\theta,P_\sigma), \ldots , (\theta_n,r_n, P_\theta,P_\sigma) > \quad (33)$$

where θ_i and r_i are the polar coordinates of each object point with respect to the object centroid. P_θ and P_σ represent the orientation and scale of the dominant curve point as defined in (32). Finding correspondences between points in each instance allows the deformation of that point with respect to the object centroid to be quantified. The first instance I_i (i=1) acts as a reference instance to which all subsequent instances are mapped. Since deformation results in the displacement, emergence and disappearance of dominant curve points between individual instances, each instance is defined by an arbitrary number of dominant curvature points and a 1:1 correspondence between these points cannot be guaranteed. To overcome these problems the binding of points between two instances is split into three phases. The first looks for salient features, or landmarks on each object, such as maxima or minima in the polar coordinate representation, as shown in figure 5.19(b). It is important that a 1:1 correspondences exists between these landmark points. For this work, we have taxonomic landmark points present on each of the objects of interest to serve as these common feature points. Since the circle sub-structures C1 and C2 represent landmark features themselves, in order to model the deformation between each training instance two imaginary landmark points are created on each boundary at $\theta_1=\pi/4$ and $\theta_2=5\pi/4$.

For each object instance I_i a landmark vector is created

$$L_i = < l_1, l_2 , .. , l_m > \quad (34)$$

where there are m landmarks for each instance, and each landmark l_i is defined as

$$l_i = (\theta_i ,r_i) \quad (35)$$

The mapping of landmark point L_j in instance i to the corresponding landmark in the reference instance is defined as

$$\begin{aligned} L_{ij}\,\theta' &= L_{1j}\,\theta \\ L_{ij}\,r' &= L_{1j}\,r \end{aligned} \tag{36}$$

The second phase takes all non-landmark points and maps them to the reference instance using the relevant landmark point mappings. The mapping of point P_j in object instance i (i>1) to the reference instance is defined as

$$P_{ij}\,\theta' = L_{1a}\,\theta + (P_{ij}\,\theta - L_{ia}\,\theta)\,\frac{L_{ib}\,\theta - L_{ia}\,\theta}{L_{1b}\,\theta - L_{1a}\,\theta} \tag{37}$$

$$P_{ij}\,r' = \begin{cases} P_{ij}\,r\,\dfrac{\lambda R}{\lambda I} & \text{if } (P_{ij}\,r \le \lambda I) \\ \lambda R + (P_{ij}\,r - \lambda I)\,\dfrac{\lambda R}{\lambda I} & \text{otherwise} \end{cases} \tag{38}$$

where

$$\begin{aligned} \lambda I &= \lambda I'\,(P_{ij}\,\theta - L_{ia}\,\theta) + L_{ia}r \\ \lambda R &= \lambda R'\,(P_{ij}\,\theta' - L_{1a}\,\theta) + L_{1a}\,r \end{aligned} \tag{39}$$

$$\begin{aligned} \lambda I' &= (\,L_{ib}r - L_{ia}r\,) \;/\; (\,L_{ib}\theta - L_{ia}\theta\,) \\ \lambda R' &= (\,L_{1b}r - L_{1a}r\,) \;/\; (\,L_{1b}\theta - L_{1a}\theta\,) \end{aligned} \tag{40}$$

The resulting mapping is only an approximate fit of instance I_i to the reference instance. The third phase performs the actual binding of the instances by linking points that minimise the error

$$E = \sqrt{P_{ij}\,\theta' - P_{1k}\,\theta \;+\; P_{ij}\,r' - P_{1k}r} \tag{41}$$

where P_{ij} is the current point in instance i to be bound and P_{1k} is the reference point it can bind to by satisfying proximity and curvature scale constraints. If P_{ij} cannot be bound to any reference point, the reference instance is augmented by interpolating the position of a new reference instance point to which P_{ij} is then bound. When all training instances are bound, each reference point P_{1k} becomes a vector of all the point instances which correspond to that

reference point

$$R_k = (I_1, I_2, I_3 , \ldots , I_n) \tag{42}$$

where I_1 is the initial reference instance for point R_k. Each point instance $I_1 \ldots I_n$ for a given reference point R_k defines a region in which R_k deforms with respect to the object centroid. The extrema of this region are defined as

$$Rext_k = (\theta_{min}, \theta_{max}, r_{min}, r_{max}) \tag{43}$$

The deformation of the object centroid with respect to each point is thus defined as

$$Rext_k' = (\theta_{min} + \pi, \theta_{max} + \pi, r_{min}, r_{max}) \tag{44}$$

Figure 5.20 shows the final deformation regions for the two landmark points on the eye sub-structure. Each dominant curve point on each sub-structure has a similarly defined deformation region. However, the final representation is over complete. Since there is no guarantee a 1:1 correspondence between all dominant points on all instances will exist due to local shape deformation, the final object stores all dominant points as illustrated in figure 5.21. This is inherently redundant because a single instance will not contain all of the information in the final object representation. However this is necessary to handle the natural deformation present in the wasp head sub-structures.

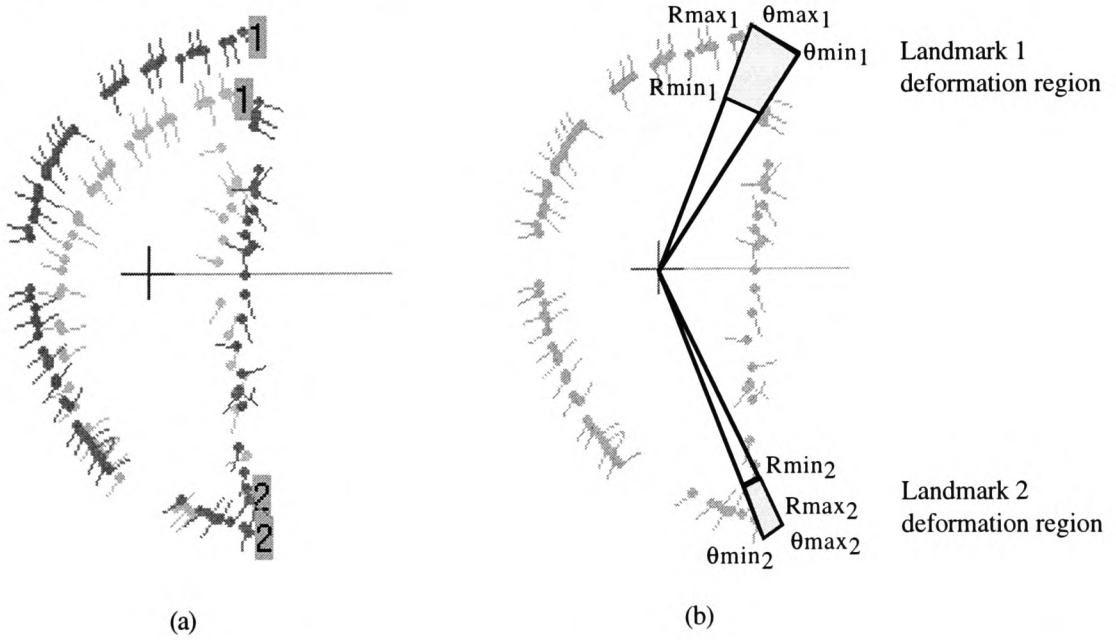


Fig 5.20. (a) Eye deformation (b) Deformation regions for the landmarks that reside on the eye sub-structure.

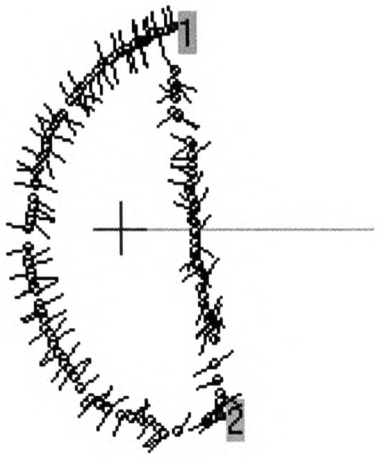


Fig 5.21. The final shape is over complete, containing all dominant curve points from the training set.

Although the above polar coordinate representation is generic, the algorithm for binding shape instances is not. It is important that the polar representations used satisfy the centroidal profile constraint. This states that only one edge point can lie on each radial line from the object centroid. Each of the wasp head sub-structures satisfy this constraint. The top of the head sub-structure causes problems in that dense regions along the orientation axis of the polar representation fluctuate rapidly along the radial axis. This is due to the close proximity of edge points on the orientation axis either side of the object centroid. This can be overcome by displacing the object's reference point so that it no longer corresponds to the object centroid. Figure 5.22 shows how moving the object's reference point eliminates problems of dense regions of points in the polar representation. This problem however does not affect the other wasp head sub-structures.

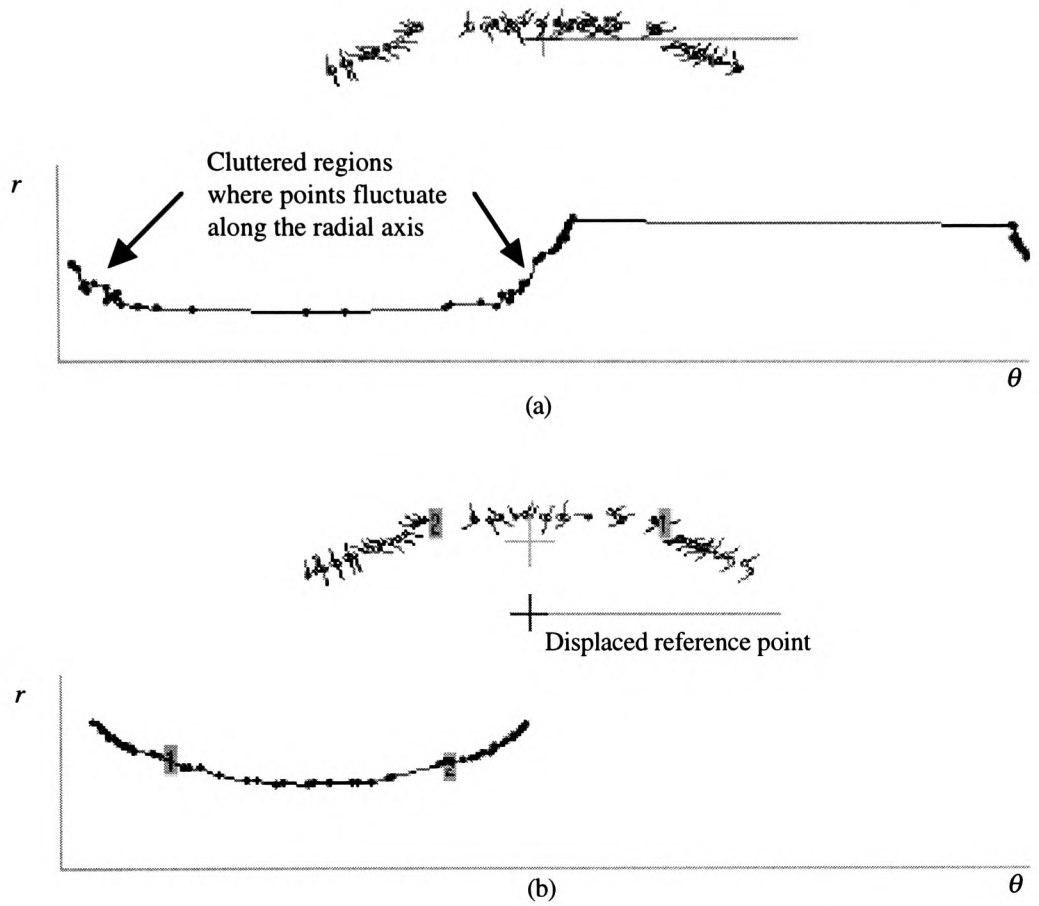


Fig 5.22. (a) The reference point equals the object centroid. Here dense regions of the polar representation fluctuate significantly along the radial axis. (b) Displacing the reference point (+20 pixels downwards) eliminates this problem resulting in a smooth curve in the polar representation.

The principles described above for modelling deformation at the sub-structure level also apply to the compound object level. However the situation is simplified compared to binding dominant curvature points on the sub-structures themselves because we want to model the deformation of sub-structure centroids with respect to the centre of gravity for the wasp head. Just as corresponding dominant curvature points between sub-structure instances define deformation regions, so the displacement of object centroids with respect to the centre of gravity define deformation regions for each of the wasp head sub-structures. Figure 5.23 shows an example of two wasp head models, illustrating the deformation between them as well as the deformation regions for each sub-structure centroid with respect to the centre of gravity of the wasp head.

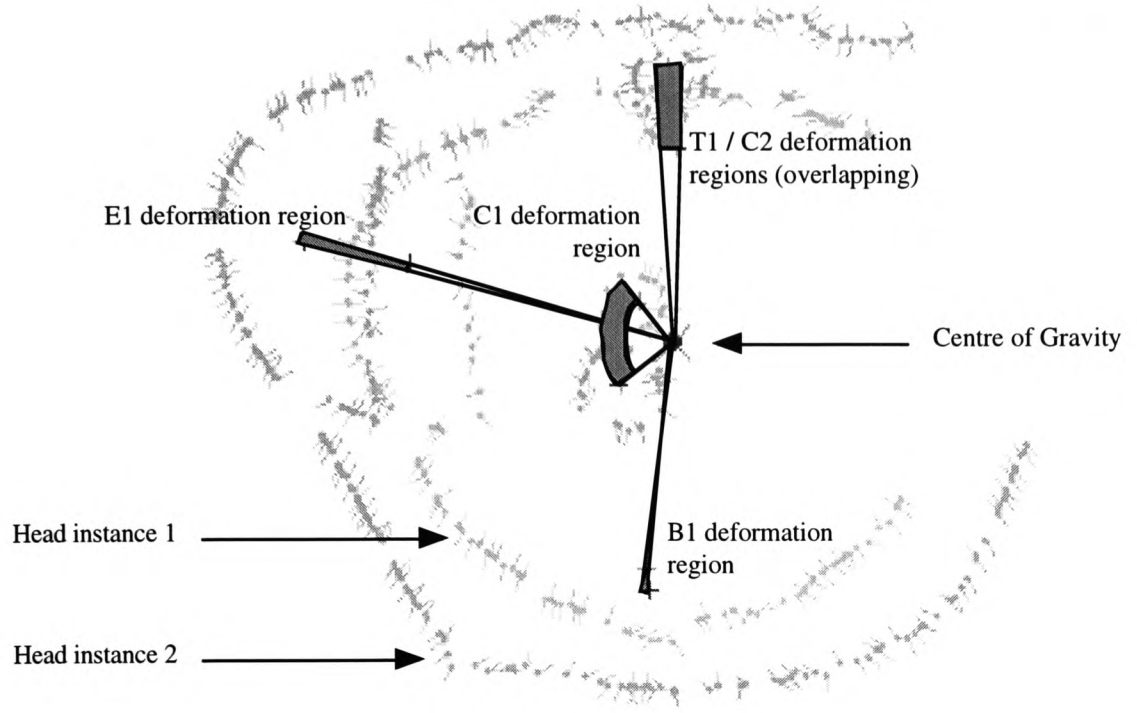


Fig 5.23. Deformation regions for the sub-structure centroids at the compound object level.

5.5. Conclusion

The problems of discontinuous contours have been discussed and a simple 1D contour model and extension methodology have been presented to overcome these problems. A novel technique which uses the statistical properties of contour segment lengths to filter out reintroduced textural features is presented. This represents part of the contribution to knowledge for this work. It is acknowledged that we cannot assume the input data to any image processing algorithm satisfies idealised conditions and that subsequent processing must be robust enough to handle variable and noisy data. Given the discontinuities between contour segments and residual textural clutter in the contour map, the use of 1D curvature representations to describe the wasp head sub-structures is deemed unsuitable. A hierarchical model based on the dominant curvature point representation is presented. This representation is based on zero-crossings so it overcomes the problems of maxima variance over curvature scale space and retains Mokhtarian and Mackworth's uniqueness, invariance and stability properties of shape representation [89]. This representation is used to model the deformation between instances of wasp heads and their respective sub-structures. A novel algorithm is presented which automatically finds correspondences between dominant curve points on deformed instances of an object and allows the deformation between those instances to be quantified. This approach does not rely on exact 1:1 correspondences required by other statistical modelling techniques such as Active Shape Models (ASMs). The dominant curvature

point representation described above serves as a better input domain for subsequent object and landmark localisation than the original image does for other deformable contour/ shape models such as snakes and Active Shape Models. This is because the textural clutter in the original wasp head image causes these deformable techniques to produce sub-optimal results. As such, this representation is important for subsequent object and landmark localisation.

Chapter 6

Object Localisation

6.1. Introduction

Chapter 5 describes how the intermediate contour map representation can be used to derive a high level dominant curvature point representation encoding the salient shape information of an object. From this representation, a hierarchical model of the wasp head and its sub-structures was described. This represented the deformation between instances of an object, or wasp head sub-structure and the wasp head as a whole. Given this representation, the next stage is to locate the actual landmarks in each wasp head image. In order to automatically locate the landmarks on each wasp head, a localised neighbourhood must be defined for each taxonomic landmark point. Segmenting the search space in this way reduces problems of ambiguity caused by correlated noise in the input image [68]. Using the shape models described in chapter 5, the deformation region of each landmark defines its neighbourhood relative to the object, or sub-structure centroid. Since the representation is translation invariant, the deformation region of each landmark point cannot be associated with any particular region of an input image. Therefore, in order to localise these landmark neighbourhoods each object centroid must be located first. This chapter focuses on the use of elastic graph matching and the generalised Hough transform to localise the eye (E1), top of head (T1) and bottom of head (B1) sub-structures. Both the "top hat" and smooth kernel versions of the Hough transform will be investigated. Since the circle objects (C1 and C2) are landmarks themselves, localisation of these will be discussed later. The results show that the smooth kernel Hough transform produces better localisation results than either the graph matching or "top hat" Hough transform techniques. This will thus be used for subsequent landmark localisation. No original contribution to knowledge is presented in this chapter as it investigates the application of deformable object localisation techniques to the wasp head data shown in appendix 1.

6.2. Elastic Graph Matching

Given the deformation present in the shape models described above, it is important that the object localisation techniques used are sufficiently robust. Since elastic graph matching can be applied to problems where the input contains significant deformation [122], it is used here to localise the objects of interest on each wasp head.

6.2.1. Experimental Design

The elastic graph matching algorithm is applied to the problem of localising the eye sub-structure within a set of test images. The eye model is created using a training set of 6 instances from the wasp head images {e100,e200,e201,e202,e204,e405}. These instances are from species 1,2 and 4, as shown in appendix 7. Instances of the eye sub-structure from

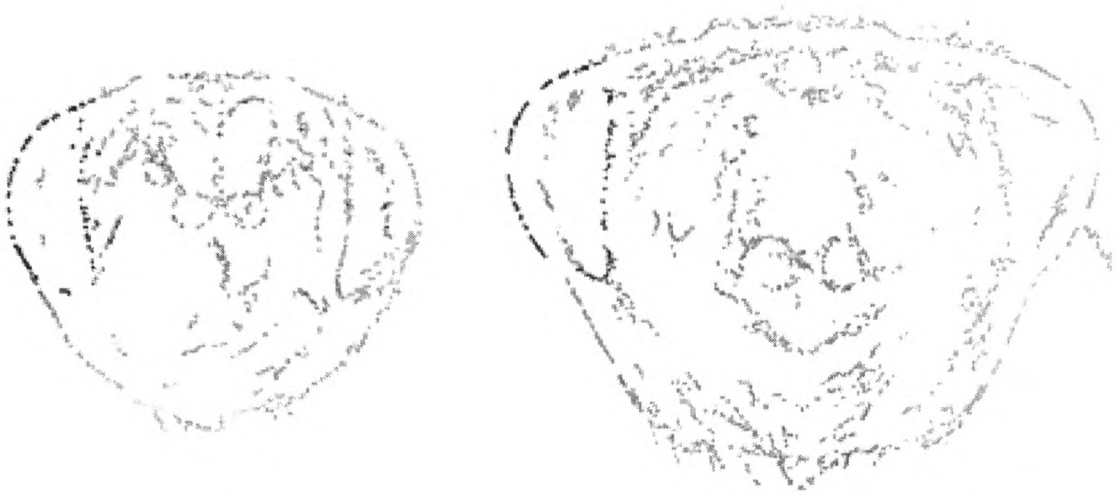
species 3 and 5 are "unseen" during training. The eye model is then applied to the test image set {e100,e200,e300,e305,e400,e501}. The latter 4 test images contain instances of the eye sub-structure not used during training so the effectiveness of the graph matching technique when applied to unseen deformation can be tested.

6.2.2 Algorithm Design

Elastic graph matching uses optimisation techniques to fit a stored model to an input image in order to find graph or sub-graph isomorphisms, where the stored model matches all or part of the input data respectively. A graph is defined by a set of nodes and arcs. Here, nodes correspond to the dominant curvature points that make up the sub-structure model and arcs correspond to the spatial relationships between them. Explanations of the basic graph matching methodology can be found in Sonka *et al.* [122]. Here the spatial error between the stored model and the corresponding points in the input dominant curve point map is minimised. The set of points which minimise this error are considered to define the object of interest in the input data. Since only one instance of each object exists in each wasp head image, problems of locating multiple instances are not relevant here.

6.2.3. Graph Matching Results

Figure 6.1 shows the results of the above algorithm when applied to test images e100 and e200. The brightness of each point in the output image represents the likelihood that they correspond with their respective points in the stored eye model, with darker points representing a good match. Since images e100 and e200 are used during training, the points around the eye sub-structures produce a good output response. However, other dominant curve points which do not correspond to the eye sub-structure also produce an output. This is because no prior assumption about the location of the eye can be made, so the graph matching algorithm analyses the entire image. Correlated image noise results in points which are extraneous to the eye sub-structure producing an output, but these false responses decay with the likelihood of being part of the eye.



e100 e200

Fig 6.1. Graph matching results for test images e100 and e200.

Figure 6.2 shows the results of the same algorithm applied to test images e300, e305, e400 and e501, all of which contain instances of the eye object unseen during training. As can be seen, the output for the points that define the eye sub-structure is significantly less than in the results for experiments e100 and e200. This is because the spatial relationships between the input points vary from those in the stored model as a result of natural deformation. It can be seen that the outer edge of the eye has a larger response than the inner edge because this part of the structure is more stable. Greater variance along the inner edge of the eye results in a weaker output for the constituent points.

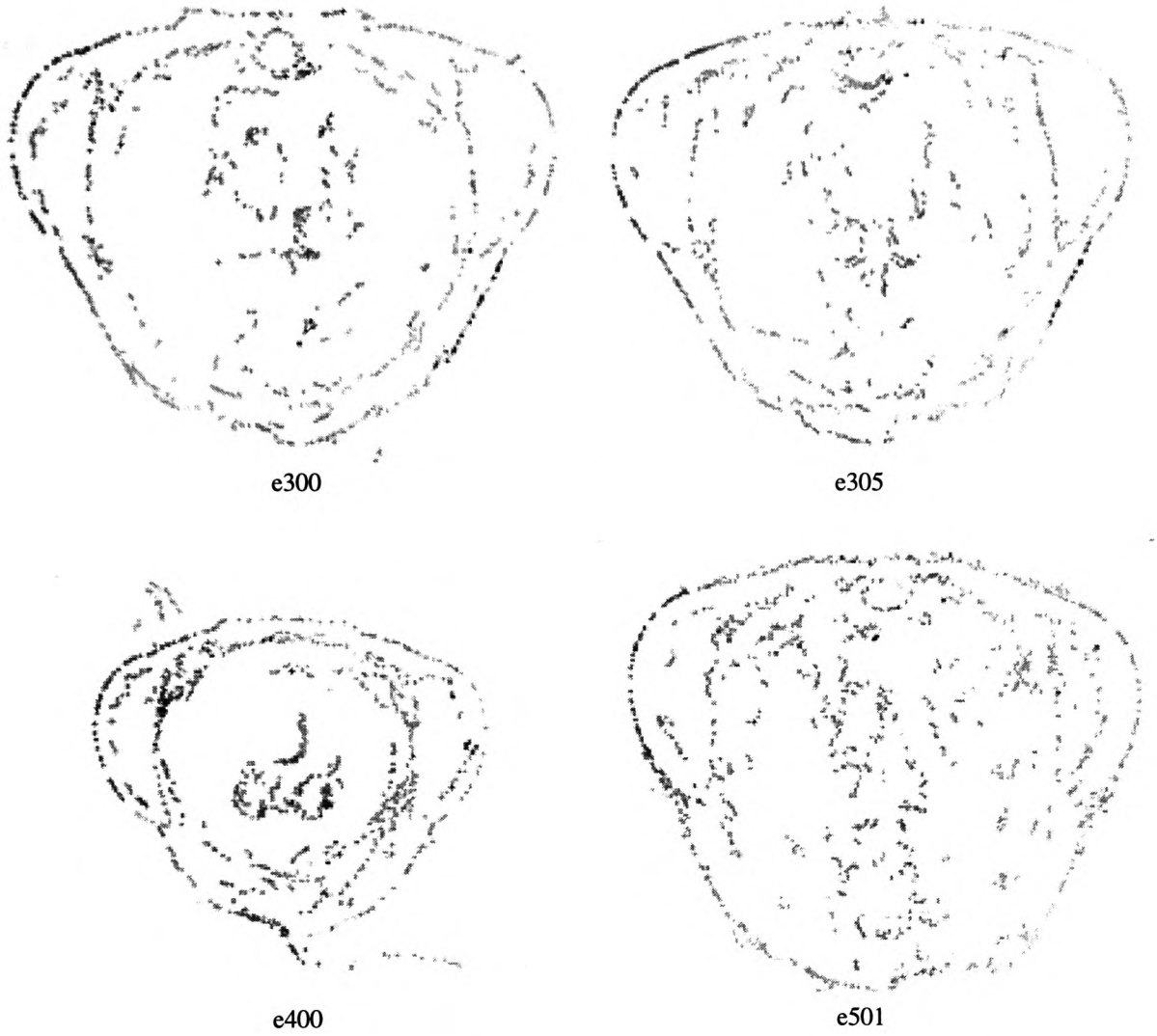


Fig 6.2. Graph matching results for test images e300, e305, e400 and e501.

It is important that sufficient deformation is contained in the training data set so the variance in the spatial relationships which give rise to the poor results in figure 6.2 can be incorporated into the stored object model. The outer edge of the eye produces a better result because the deformation is less significant, and as such can be encoded with fewer training instances. Computationally the graph matching algorithm is expensive, requiring an average execution time of 120 seconds on a 300 MHz Pentium II to produce the final output for each input image. This is because no assumption can be made about the location of the object in the image and input points extraneous to the eye must also be considered. Also, the graph matching methodology matches all points in the stored model to the input dominant curve point map. Since we only want to localise the taxonomic landmark points, a more efficient algorithm would be to locate the object centroid (only one point) from which we could then find the actual landmarks.

6.3. The Conventional Generalised Hough Transform

The generalised Hough transform [7] serves as an extension to Hough's original algorithm [47] to locate objects which cannot be described analytically. Since this algorithm has been widely used throughout the literature, it is applied here to determine its effectiveness in localising the objects of interest on the wasp head. The conventional Hough transform uses a "top hat" voting kernel, as illustrated in figure 6.3, to modify the parameter space, or accumulator array [93].

6.3.1. Experimental Design

To determine the effectiveness of the conventional Hough transform, these experiments again focus on locating the eye sub-structure. As above, a model of the eye object is derived from the set of training images {e100,e200,e201,e202,e204,e405}. The conventional Hough transform is calculated for test images e200, e300 and e501. As above, e300 and e501 contain deformation unseen during training. Since the focus of this work is to localise objects under conditions of natural deformation, global rotation and scaling are not considered here. As as result the parameter space has the same dimensionality as the input image. The basic mechanics of the generalised Hough transform are discussed in detail by Ballard *et al.* [7]. The parameter space is modified with a simple "top hat" function usually configured to be 1 pixel wide [92].

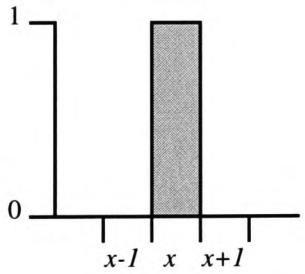


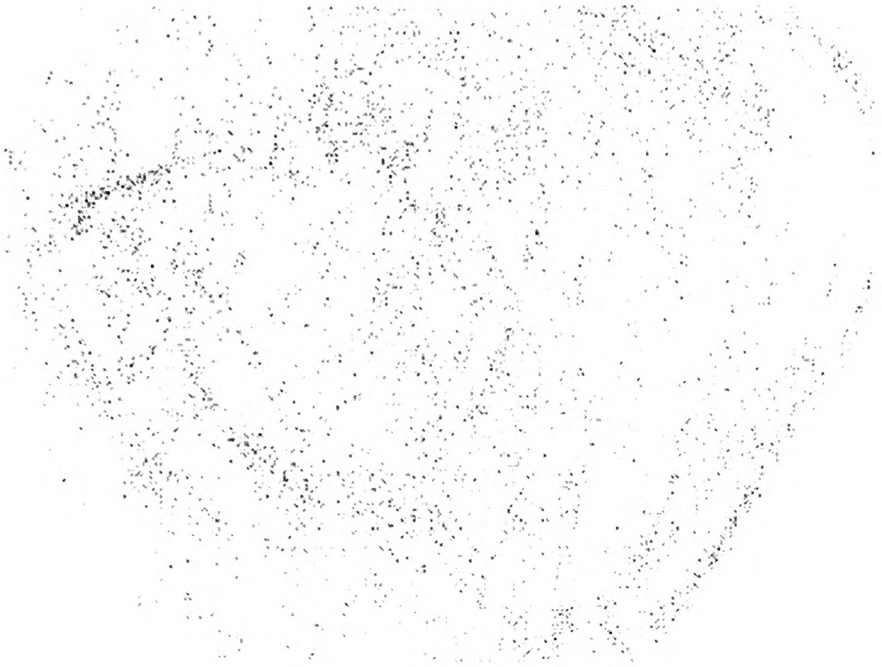
Fig 6.3. Top hat function used by the Hough transform.

6.3.2. Results

Figure 6.4 shows the results of the conventional Hough transform applied to test images e200, e300 and e501. Given that the dimensionality of the parameter space is the same as the input image, the time required to calculate the Hough transform is significantly less than the time required to perform the above graph matching algorithm, with an average execution time of 0.5

seconds. Each point / maximum in the resulting parameter space represents a possible object (centroid) location. Since image e200 is used during training, a single maximum exists in the resulting parameter space, as shown in figure 6.4(a). Hence a good localisation of the object centroid is obtained. Since images e300 and e501 contain unseen deformation, the resulting localisation is much poorer. A significant number of false maxima exist not only around the location of the object centroid, but also over the entire parameter space. This can be seen in figures 6.4(b) and 6.4(c). The dispersal of maxima points results firstly from object variation. Since the location of dominant curve points can change as a result of local object deformation, the relative position of the object centroid also changes. Hence many points in the parameter space are modified with the "top hat" kernel, resulting in a large number of false object locations. The second major problem is correlated noise. As with the graph matching results, points extraneous to the eye object must also be considered since no prior assumptions can be made about its location. Figures 6.4(b) and 6.4(c) clearly show the false maxima produced for test images e300 and e501 respectively. Correlated noise is also present in the output for experiment e200, shown in figure 6.4(a). However, since the results are normalised into the range [0,1], the resulting image does not highlight false maxima because their values are negligible. Figure 6.5 shows the same result for experiment e200 but the display transfer function used is logarithmic, not linear. It is clear that correlated noise still has an effect on the results.

(a)



(b)

(c)

Fig 6.4. Conventional Hough transform of the eye object for experimental images (a) e200, (b) e300 and (c) e501. Good localisation is achieved for image e200 but poor localisation results are obtained for images e300 and e501 due to deformation.

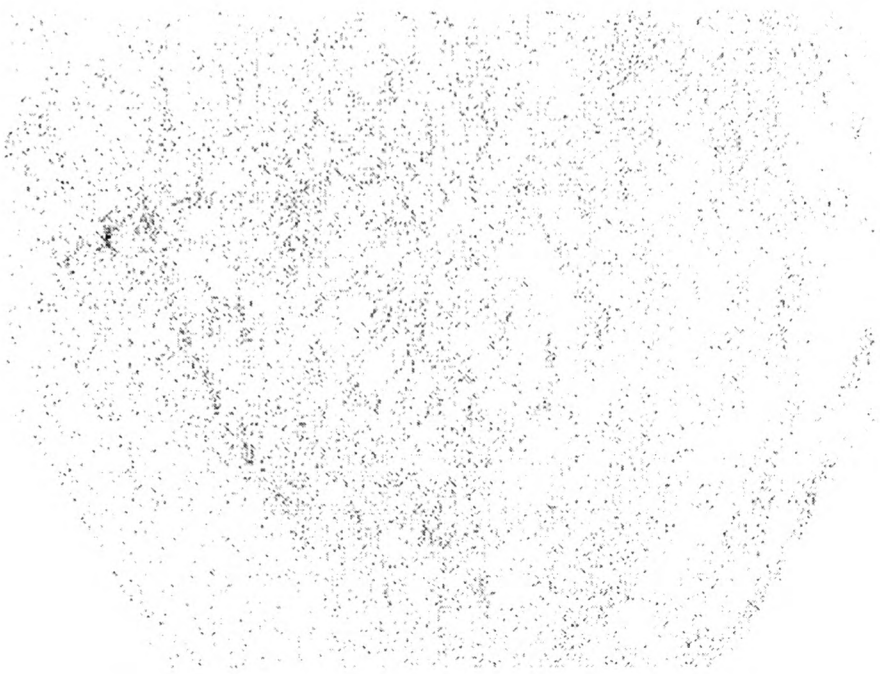


Fig 6.5. Hough transform for experimental image e200 shown using a logarithmic display function.

High level knowledge can be used to guide the search for the true object location. Figure 6.6(a) shows the compound object representation for the training image set used above. The resulting deformation region for the eye object is shown in figure 6.6(b), overlaid on the Hough transform for image e501. Within this deformation region, a significant number of maxima exist which correspond to false object (centroid) locations. As noted above, the deformation region of each landmark point defines the neighbourhood in which the point resides with respect to the object centroid. Given the spatial validity of each maximum in the eye's deformation region, searching for each landmark in the respective neighbourhood offset from each candidate maximum would result in a significantly large search area. It is important to minimise this neighbourhood with the aim of reducing structural ambiguity for subsequent landmark localisation.

In conclusion, the conventional Hough transform performs well for input where there is little or no deformation, but poor object localisation is obtained when deformation is present. Hence it is deemed unsuitable for locating the wasp head sub-structures and for calculating a sufficiently small neighbourhood in which to search for taxonomic landmark points.

(a)

(b)

Fig 6.6. (a) The wasp head compound object representation defines the deformation region of each sub-structure relative to the centre of gravity. (b) Significant numbers of candidate maxima exist in the eye's deformation region.

6.4. Smooth Kernel Hough Transform

The limitations of the conventional Hough transform described above arise from the fact that that the highly localised "top hat" function increments only one point (pixel) in the parameter space, resulting in poor correlation of the object's centroid when deformation is present. This can be overcome by using a smooth kernel function which is distributed over many points within the parameter space. Leavers [68] discusses the equivalence of using a smooth kernel with the maximum likelihood statistical function while Palmer *et al.* [93] show that smooth kernels produce better results when locating analytical shapes in the presence of gaussian noise. Here we want to use the deformation data at each dominant curve point to configure the smooth kernel function and apply this to the problem of locating the sub-structures on the wasp head.

6.4.1. Experimental Design

To see if the smooth kernel approach overcomes the problems of false maxima inherent in the conventional Hough transform, the experiment described in section 6.3.1 is used. The only difference is in the implementation. Here each dominant curve point defines a deformation region which is used to configure a gaussian based function, that is, the smooth kernel used to modify the parameter space.

6.4.2. Smooth Kernel Hough Transform Algorithm Design

Once the deformation properties of each point have been calculated, the next stage is to create the Rtable [7]. This is used to index each point's deformation region based on its orientation in the image plane. Since each reference point R_k defines a set of n instances, as described in chapter 5, the Rtable could contain n separate indexes to R_k. A more effective way would be to cluster instances of R_k according to their orientation. To do this, R_k as defined in (42) is ordered according to the orientation of each instance in the image plane.

$$R_k = < I_1, I_2, I_3 , \ldots , I_n > \tag{45}$$

From this the orientation variance between each point is calculated.

$$\delta\theta_i = I_{i+1}\theta - I_i\theta \tag{46}$$

Point instances I_i and I_{i+1} are considered to belong in the same cluster if $\delta\theta_i < T$, where T is an orientation threshold which determines when two point instances belong in the same cluster.

Each cluster c_i is then defined by the orientation extrema of the point instances that define the cluster.

$$c_i = (\theta_{min}, \theta_{max}) \tag{47}$$

Figure 6.7 illustrates how point instances form the basis of clusters. The number of clusters depends on the behaviour of each point instance. If a single cluster exists then either all instances have the same orientation, the orientation variance between all instances is less that T or R_k is affected by severe deformation so the orientation of each instance is distributed throughout the $[0,2\pi)$ range. It is more likely two or three clusters will be extracted as a result of concavity / convexity changes in the boundary which results from local object deformation.

Given a set of clusters for each point, the Rtable index can be further reduced by overlaying each cluster, and deriving new clusters where existing ones overlap. This forms the final Rtable structure required for the smooth kernel Hough transform. The new Rtable eliminates the redundancy inherent in storing an index to different points for the same orientation range and the need to search the whole index for points who may have instances that lie within a given orientation range.

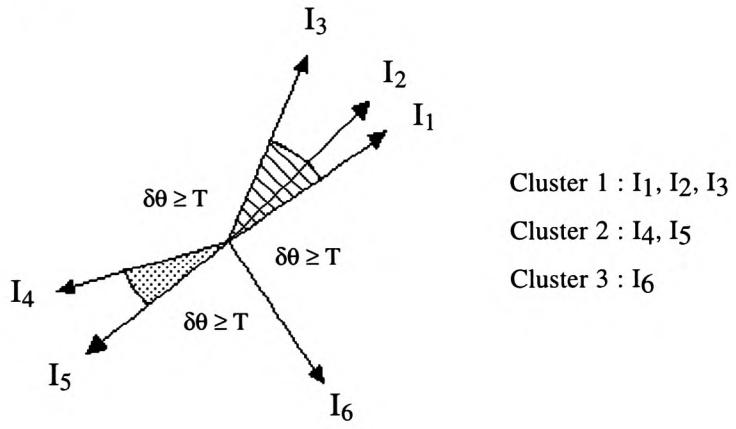


Fig 6.7. The formation of clusters based on the direction of each point instance and the orientation variance between each instance.

The mechanics of the smooth kernel Hough Transform remain similar to the conventional Hough transform as described above. For a given input point, the orientation range it lies in is found in the Rtable index. From there, the relevant object points are accessed. For each point, the deformation region of the object centroid, as defined in (44), can be addressed. In the conventional Hough transform the centroid is mapped to the appropriate point in the parameter space relative to the current input point. For the smooth kernel Hough, the deformation region is mapped to an appropriate region of the parameter space. Instead of incrementing a single

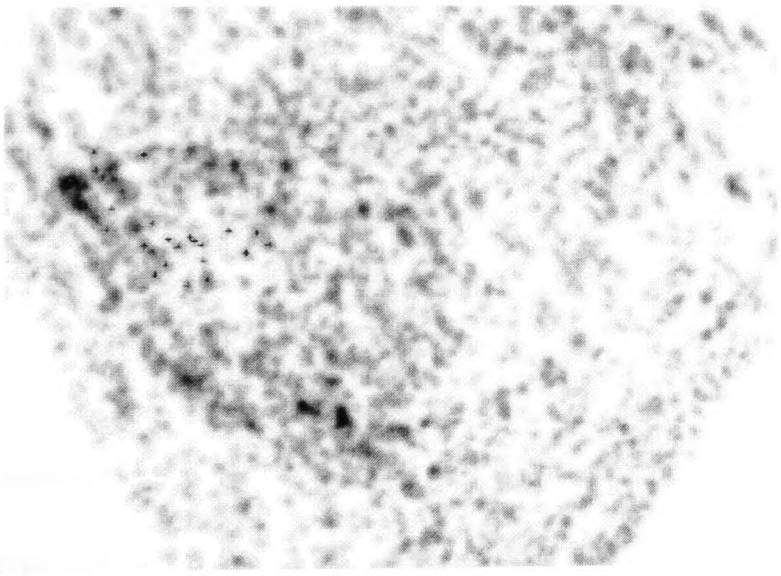
accumulator array element, all elements that fall within the deformation region are incremented. These are modified according to circular gaussian mapped into the deformation region. The increment in each accumulator array element that falls within the deformation region is defined as

$$\Delta A_{xy} = 1 + \frac{-2}{1 + e^{-\sqrt{x'^2 + y'^2}} - 1} \tag{48}$$

and x',y' are the coordinates of each accumulator array element A_{xy} mapped into the circular gaussian space where x'= A_x=>[-1,1] and y'= A_y=>[-1,1]. The mapping of deformation regions to the parameter space instead of individual points is considered important in handling object deformation.

6.4.3. Results

Figure 6.8 shows the resulting parameter space for test images e200, e300 and e501 after applying the smooth kernel Hough transform. Darker regions correspond to possible locations of the eye sub-structure.



(a)

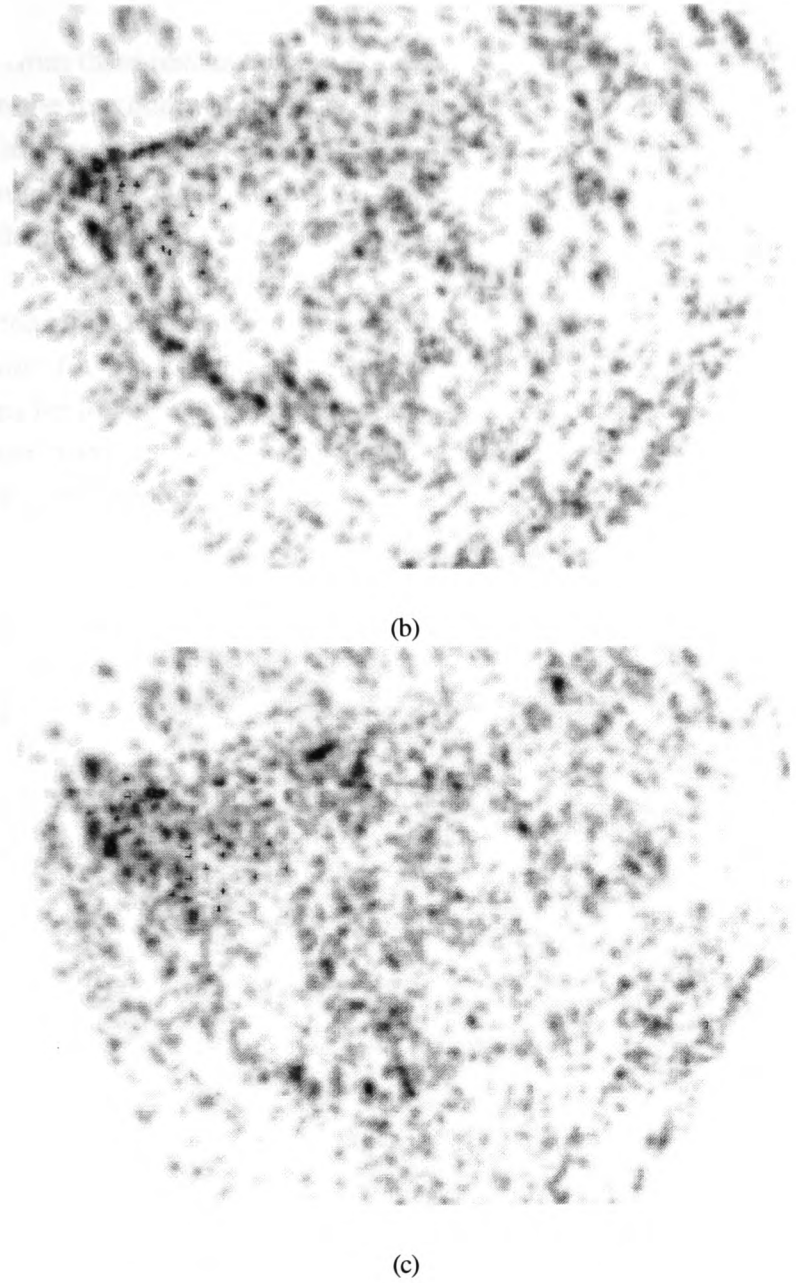


Fig 6.8. Smooth kernel Hough transform results for images (a) e200, (b) e300 and (c) e501.

From the above results, a significant difference that is evident between the smooth kernel and conventional transforms is that localised objects are represented as regions in the parameter space, not individual points. The shape of these regions reflects the deformation between instances of a given object. As noted above, individual points in the conventional transform are unlikely to correlate in the parameter space, giving rise to large numbers of disparate points, or false object locations, as shown in figure 6.4. In the smooth kernel approach, deformation regions are more likely to overlap, giving rise to points with higher values (shown as darker regions) at possible object locations.

It is also evident from these results that the problem of false regions having maxima equal to or even greater than the maximum at the true location of the object remains. Grimson *et al.* [43] note that such false responses are an unavoidable effect of using techniques based on the generalised Hough transform. However, another significant difference between the conventional and smooth kernel Hough approaches is that the number of false object locations produced using the smooth kernel approach is significantly less than the large number of dispersed maxima produced using the conventional approach. This can be seen in the localisation results for the bottom of the head object. Figure 6.9(a) and 6.9(b) show the Hough transforms for image e300 using the "top hat" and smooth kernels respectively. Clearly the number of false maxima within the resulting parameter space is significantly reduced when using the smooth kernel approach.

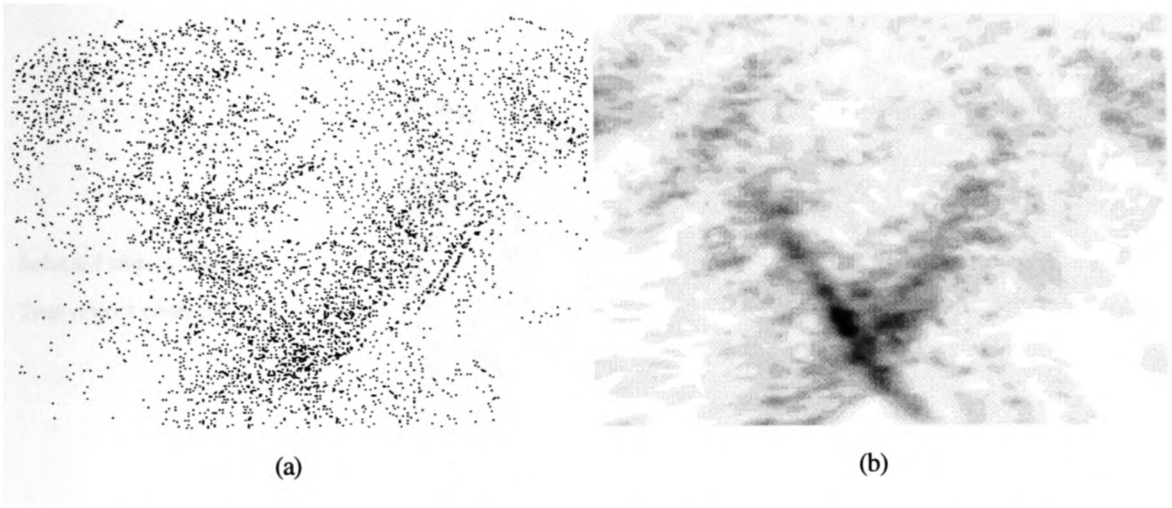


Fig 6.9. (a) Conventional "top hat" Hough transform. (b) Smooth kernel Hough transform.

This is an important result as it shows that, unlike the conventional Hough transform, the problem of false maxima, or false object locations produced within the object's deformation region does not necessarily worsen under conditions of local deformation and clutter. As noted above, the maxima in the deformation region of a given object are equally valid and no one maximum can be assigned precedence over the others. Hence all maxima in the deformation region of an object centroid must be considered. It is therefore the minimisation of false maxima and the resulting reduction in the size of the landmark neighbourhood which is important here. Given that fewer false maxima exist in the object's deformation region when using the smooth kernel Hough transform, the size of the resulting landmark neighbourhood is reduced, thus overcoming the problems inherent in the conventional Hough transform. Also, as a result of the deformation present in the input dominant curve point map and encoded in the

stored object models, the maximum which corresponds to the location of the object centroid is actually slightly offset from the true centroid location. This is illustrated in figure 6.10. However, the offset is only slight and as such the effect on the localisation of the landmark neighbourhood is negligible. Finally, the processing requirements for the smooth kernel Hough are greater than the conventional transform because more points in the accumulator array are modified. An average execution time of 20 seconds was obtained for each input image.

In conclusion, the smooth kernel Hough transform overcomes the limitations of the conventional "top hat" approach by significantly reducing the number of candidate object locations. This has the effect of reducing the size of the neighbourhood in which to search for taxonomic landmark points. Due to deformation, the maximum which corresponds to the object's centroid is slightly offset from the true centroid location. However, this is not significant and the resulting displacement of the landmark neighbourhood is negligible.

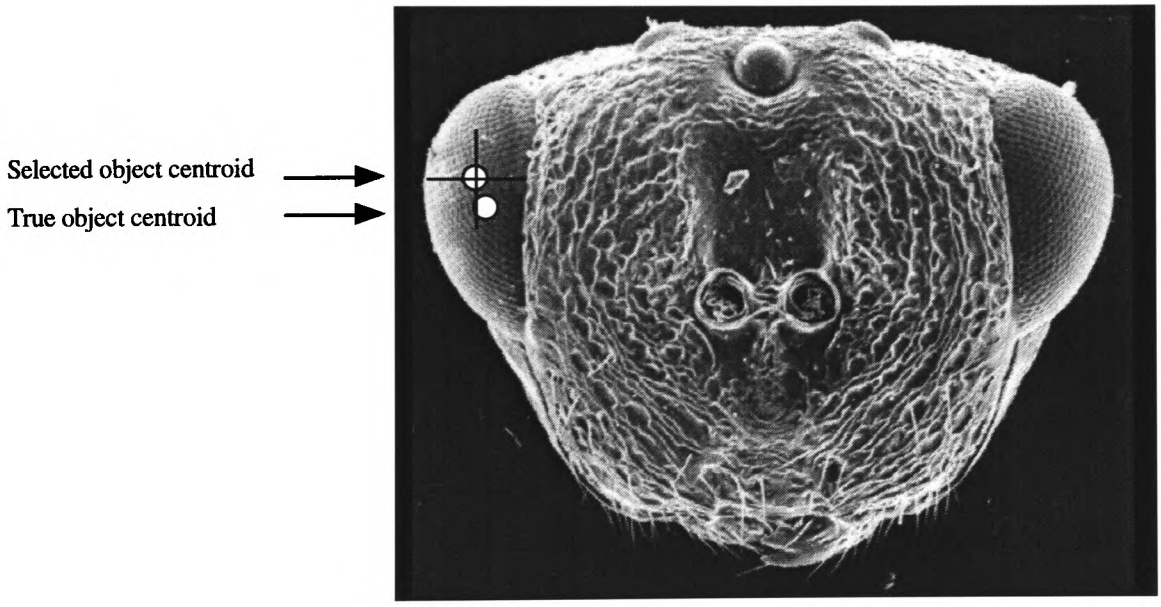


Fig 6.10. Deformation results in the displacement of maxima that correspond to an object's centroid.

6.5. Encoding Deformation

The experiments carried out above have been using the training set illustrated in appendix 7. It is important to determine how well the smooth kernel Hough transform performs when different training data is used. To investigate this, three different training sets are used to localise the eye object in test images e100, e200, e300, e305, e400 and e501. The training sets are defined as {e100,e101}, {e100,e200} and {e100,e200,e400}. Figures 6.11, 6.12 and 6.13 show the smooth kernel Hough transform results for the eye object using training sets {e100,e101}, {e100,200} and {e100,e200,e400} respectively. As above, dark regions in the parameter space correspond to candidate object locations. As the number of training instances increases, so the localisation of the eye object improves, with significant maxima emerging around the location of the eye centroid. The reason for this is illustrated in figure 6.14 which shows the different configurations of the voting kernel for a single point on the eye object for each of the training sets used in this experiment. As the number of training instances increases so the kernel used to modify the accumulator array becomes larger. This is important because the results discussed above for the conventional "top hat" and smooth kernel Hough transforms show that in the presence of object deformation, smaller voting kernels are less likely to correlate in the parameter space. This results in many false maxima being created, making object and subsequent landmark localisation difficult. As the size of the training set decreases, so does the size of the voting kernel, reducing to fixed template matching when only one instance of the object is stored. Since each update function is based on the deformation extrema of each point, it is sufficient that the training set contains this deformation extrema.

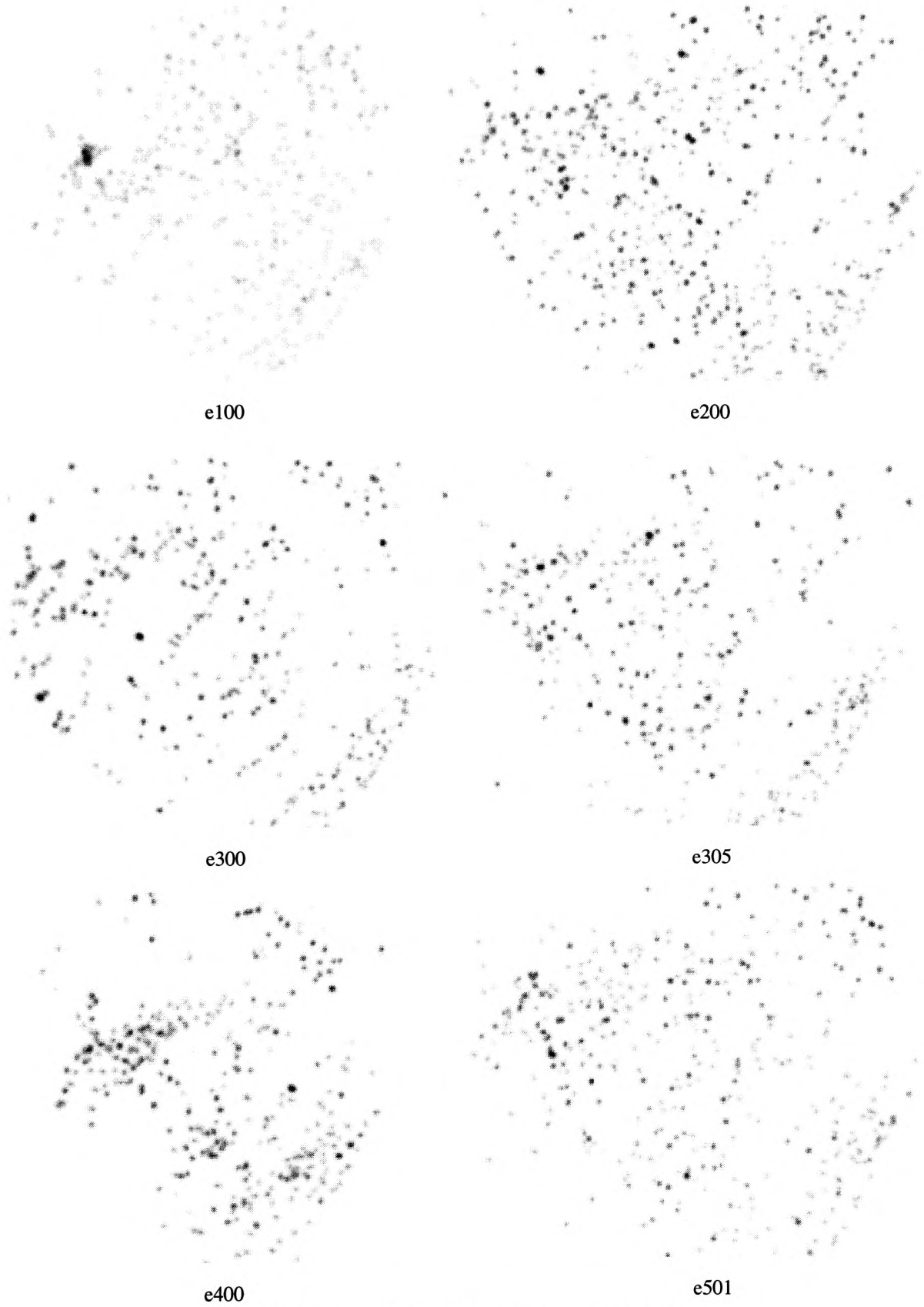


Fig 6.11. Eye localisation using training set {e100,e101}

Fig 6.12. Eye localisation using training set {e100,e200}

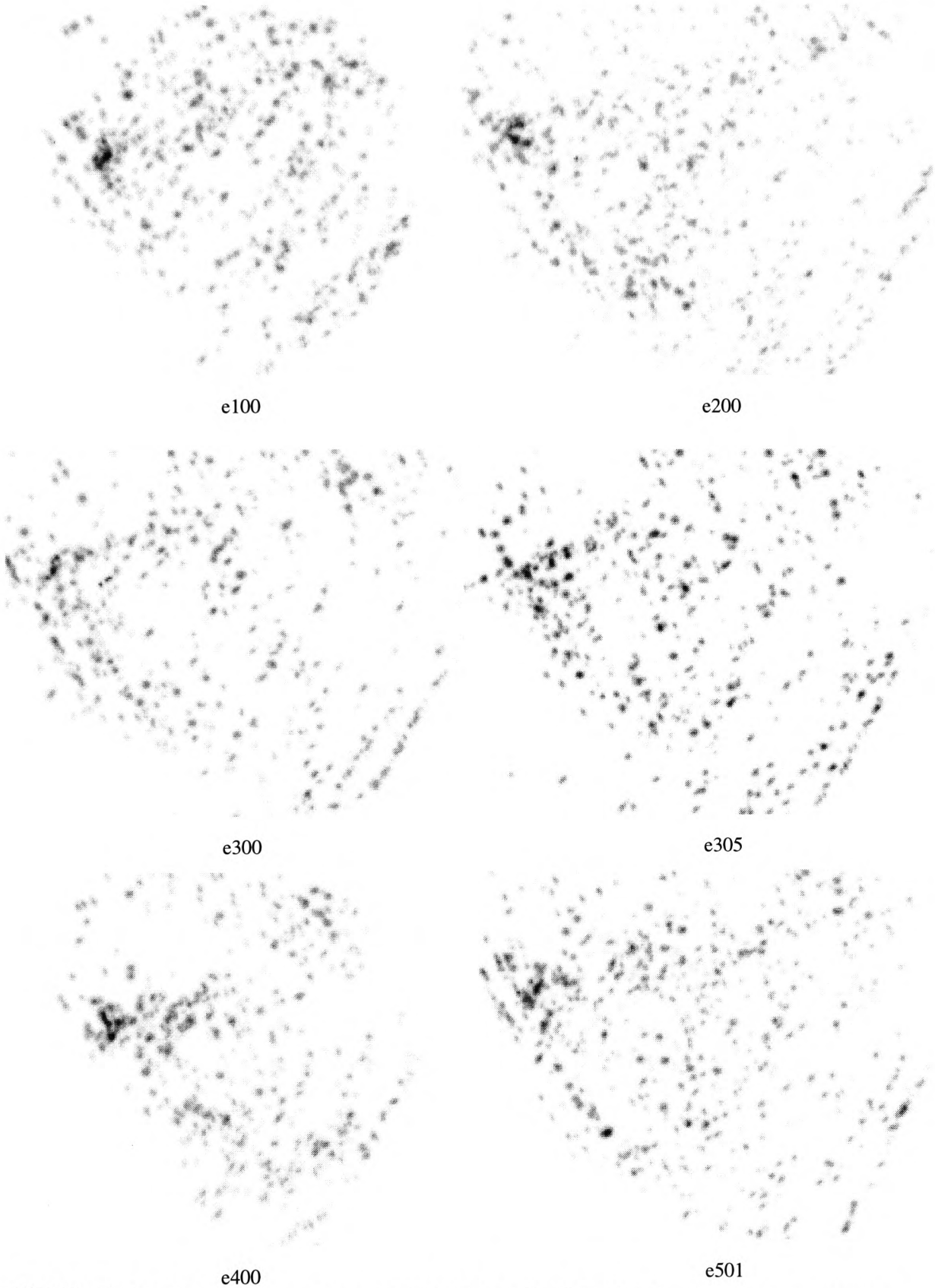
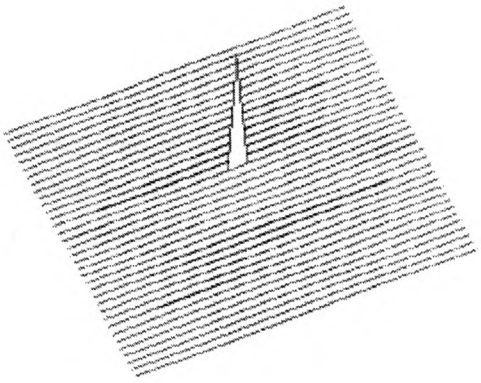
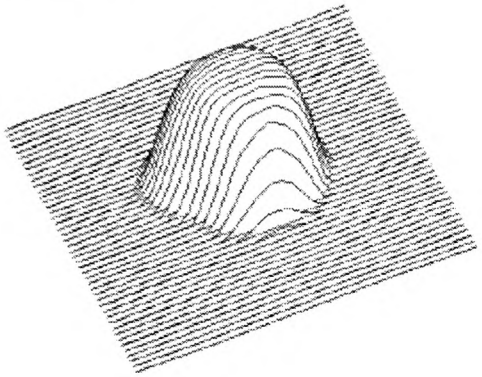


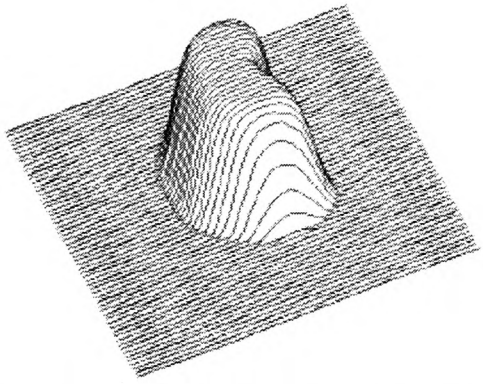
Fig 6.13. Eye localisation using training set {e100,e200,e400}. Object locations become more prominent.



Hough voting kernel for training set {e100,e101}. Poor localisation is obtained using this function, as shown in figure 6.11.



Hough voting kernel for training set {e100,e200}.



Hough voting kernel for training set {e100,e200,e400}. Better results are obtained using this function, as illustrated in figure 6.13, because more deformation is encoded in the stored object model.

Fig 6.14. Voting kernels used to modify the Hough parameter space.

6.6. Conclusion

Both elastic graph matching and the generalised Hough transform have been applied to the problem of localising the wasp head sub-structures. Although the graph matching results were good where little object variation was present, the time taken to fit the stored model to the input dominant curve point data was prohibitive. Both the conventional "top hat" and smooth kernel variants of the generalised Hough transform were considered. The conventional approach produced a significant number of false object locations while the smooth kernel approach produced fewer false locations in the parameter space. This reduces the size of the neighbourhood in which the respective taxonomic landmark points can reside. The smooth kernel Hough transform will thus be used to locate each sub-structure centroid for subsequent landmark localisation.

The results discussed above show that it is important the deformation extrema of each object point are encoded in the stored model. The deformation extrema can either be contained in the actual training images or can be extrapolated from the deformation properties of each dominant curvature point. Hence it is sufficient that only the training instances which encode the deformation extrema of an object be used. The training set described in appendix 7 and used in the above object localisation experiments is used for this work because both small scale (localised) deformations and global transformations such as rotation are represented.

Chapter 7

Taxonomic Landmark Localisation

7.1. Introduction

The structure of the wasp head as described in appendix 3 considers the taxonomic landmark points needed for species identification to be one of two types. These are landmarks that represent whole structures within the image (circle features C1 and C2) and landmarks that correspond to points on the wasp head sub-structures. The latter are located on the eye (E1), top of head (T1) and bottom of head (B1) sub-structures shown in appendix 3. Chapter 6 describes how these sub-structures are located prior to their respective landmark features in order to limit the landmark search to a localised neighbourhood in the image plane. Given the localised neighbourhood for each landmark which is offset from the respective sub-structure centroid, the next stage is to accurately locate each landmark point. First the localisation of the circle landmark features is considered. The smooth kernel Hough transform as described in chapter 6 is applied to both the dominant curvature point map and wavelet data. The results show that the wavelet data produces more accurate results and will therefore be used for the localisation of the circle landmark features. The landmarks on the eye, top of head and bottom of head are then localised. Statistical techniques are used to match the stored deformable model to the input dominant curve point data in the localised landmark neighbourhood. The features used include the local shape information and wavelet response around the landmark point, the invariant spatial relationships between landmark points and symmetry with respect to the centre of gravity. The results show that these features do not eliminate false landmark points in all cases and that the appropriate filtering of texture using the techniques described in chapter 4 can improve landmark localisation accuracy significantly. Hence the parameters used to configure the texture filter ROI curve are just as important for landmark accuracy as the statistical features described above. These findings, in conjunction with the novel technique described in chapter 4 represent the main contribution to knowledge for this work.

7.2. Circle Landmark Localisation

The circle landmarks (C1 and C2 in appendix 2) correspond to individual objects within the wasp head model. It is considered that the smooth kernel Hough transform used to locate the other wasp head sub-structures be applied to the problem of localising these features. For this experiment, the training set {e100,e200,e201,e202,e204,e405} used in the above object localisation experiments is applied. The smooth kernel Hough transform is then used to localise both of the circle features in test images e300 and e501 because they contain object deformation unseen during training. The resulting transforms for each test image are shown in figures 7.1 and 7.2 respectively.

Landmark C1 Hough transform

Landmark C2 Hough transform

Fig 7.1. Smooth kernel Hough transforms for the circle features in image e300.

Landmark C1 Hough transform

Landmark C2 Hough transform

Fig 7.2. Smooth kernel Hough transforms for the circle features in image e501.

The actual location of each circle centroid is shown with a small cross in the above images. Unfortunately, the results for landmark C1 show no significant maximum at the true location of the object. Better results are obtained for circle feature C2 however, where a significant maximum exists at the location of the object in both images. These poor results are due to the problem of contour erosion described in chapter 4. It was shown that during contour / texture delineation, circle features defined by two concentric circles whose edges lie in close proximity will be eroded in the final contour map. Figure 7.3 shows the dominant curvature point maps for images e300 and e501. It is clear that a significant part of landmark C1 has been eroded in both cases, resulting in the poor Hough transform results shown above. The smooth kernel Hough transform used to localise the other wasp head sub-structures is therefore not suitable for localising the circle landmark features when the dominant curve point map is used as input.

(a) (b)

Fig 7.3. Dominant curve point maps for images (a) e300 and (b) e501. Circle feature C1 has been eroded.

7.3. Circle Landmark Localisation using the Wavelet Transform

The poor localisation results described above are the result of contour erosion which occurs during the filtering of textural clutter. Given that the Mallat wavelet transform contains more complete edge information, this can be used in conjunction with the smooth kernel Hough transform locate each circle landmark. From the wavelet transform results shown in appendix 4, it is clear that significant clutter resides at smaller scales but the circle landmarks still exist at the largest scale of analysis $\sigma=2^3$. Since we are looking for basic circular structures, which can be described analytically, a simplified form of the smooth kernel Hough transform described above can be used. To determine if any improvement on the results shown in figures 7.1 and 7.2 can be obtained, this new approach is applied to test images e200, e300, e400 and e501.

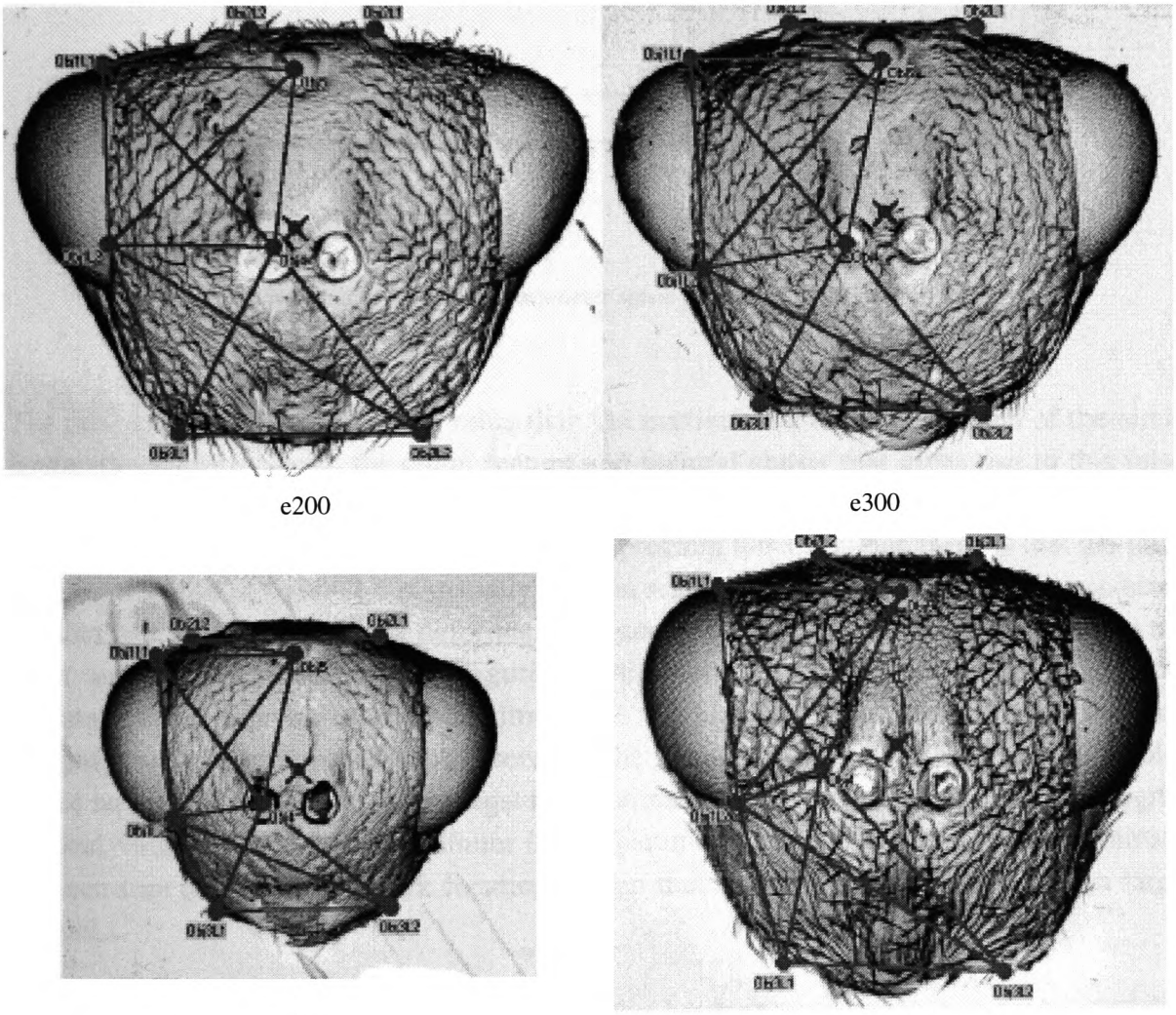
7.3.1. Algorithm Design

For each point, or pixel in a given deformation region, a one dimensional accumulator array is stored. Each point in this array corresponds to a particular radius. The training data set is used to determine a suitable range for the radius parameter. For each edge point in the landmarks's deformation region, the accumulator array of each surrounding point that lies within the pre-defined radius range with respect to the current edge point is updated. For a given point, its radius with respect to the current edge point is calculated and all of the elements in its accumulator array are updated. Each element is incremented according to a gaussian function centred on the element corresponding to the calculated radius. The use of the gaussian update function allows neighbouring elements in the array to be updated. This allows more accurate localisation and robust handling of noise as discussed by Palmer *et al.* [93] and allows the radius of the circle to fluctuate as a result of object deformation. Once all edge points have

been processed, the maximum in each accumulator array is selected as being the dominant radius at the respective point. These maxima form the final 2D parameter space from which the landmark's location is calculated. This is done by again finding local maxima in the parameter space. Where more than one maximum exists, the spatial relationships to other features on the wasp head are used. For the localisation of the circle landmarks, ambiguity is resolved by using the spatial orientation between the two circle features.

7.3.2. Results

Figure 7.4 shows the localisation results for circle landmarks C1 and C2 in test images e200, e300, e400 and e501. The landmark points are represented as dots and the spatial relationships that define salient taxonomic features on the wasp head are shown as connecting lines between each dot.



e200 e300

e400 e501

Fig 7.4. Circle landmark results for test images e200, e300, e400 and e501.

The localisation of each circle landmark is good in all of the above cases except for image e501. Here circle landmark C1 is offset from its true location. There are two main problems which result in this displacement. The first is inherent in using the wavelet transform to locate each landmark. Despite the fact that more edge information is contained in the multi-scale representation, textural clutter is also present, especially at larger scales of analysis. This results in significant false responses being produced in the Hough transform. Figure 7.5 shows the resulting parameter space for circle landmark C1 in image e501.

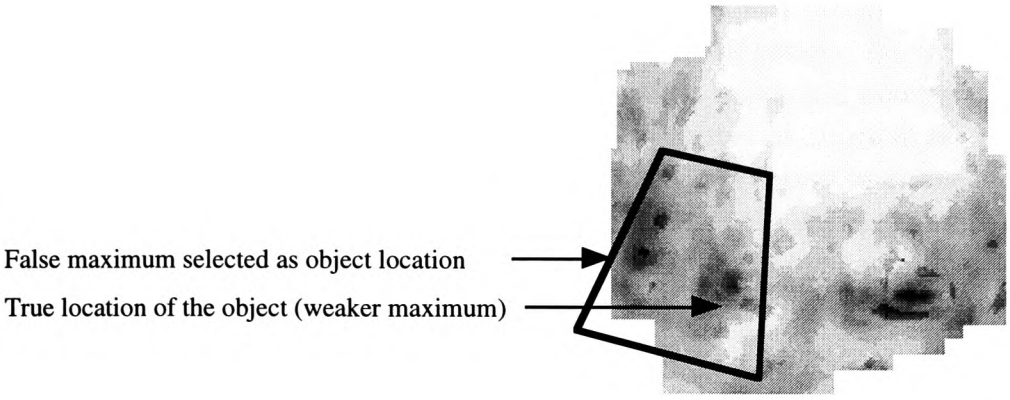


Fig 7.5. False maxima in the Hough parameter space can be selected as the object location.

The false maximum has a greater value than the maximum at the true location of the circle landmark. Unfortunately, the circle feature and textural clutter that gives rise to this false maximum are inseparable in scale space, so extending the scale range of the wavelet transform would not overcome this problem. The second problem this technique faces is that the false maximum shown in figure 7.5 is spatially valid in relation to other features on the wasp head. In other words, it lies within the deformation region of circle landmark C1 with respect to the other wasp head sub-structures. Figure 7.5 highlights the deformation region of circle landmark C1 in relation to circle landmark C2. Unlike the features used for human face recognition, the spatial relationships between the features on the wasp head are not stable. Hence no precedence can be set amongst the maxima in the deformation region. The integral, or local volume under each maximum in the parameter space also does not discriminate between true and false landmark locations given that they can both be defined by a large integral.

Unfortunately this problem is non-trivial. Since the dominant curve point map contains eroded data for circle feature C1, the wavelet transform must be used because this contains more

complete edge information. Using the edge map at the largest scale of analysis ignores the significant amount of clutter which is present at smaller scales, but the wasp head images also contain a significant amount of textural clutter at larger scales of analysis. This clutter is inseparable from the circle features so the emergence of false maxima is unavoidable. Eliminating false maxima is also difficult given that the spatial relationships between features on the wasp head are not stable. As such no priority can be assigned to any particular maximum in the respective landmark deformation region.

Appendix 8 shows the complete set of circle landmarks found for the wasp head images in appendix 1. These are located using the above smooth kernel Hough approach. To calculate the accuracy of the above algorithm, the criteria used is based on whether the location of the landmark corresponds with the location of the structure in the image. As was discussed in chapter 6, object deformation results in the offset of the maximum corresponding to its centroid. Therefore we cannot realistically expect the calculated landmark to always be in the centre of each circle feature. A landmark can thus be classed as correct or erroneous, where it lies on or outside the structure of the circle respectively. The accuracy is therefore defined as the inverse of the sum of erroneous landmarks with respect to the total number of landmarks calculated. The accuracy of landmark point j on object o, A_{oj}, is defined as

$$A_{oj} = 1 - \frac{1}{n} \sum_{i=1}^{i \leq n} E_i \qquad (49)$$

where n equals the number of instances of landmark j and E_i represents the error for landmark instance i.. Since the circle features correspond to landmarks themselves, they only have one landmark point. The above accuracy function is calculated for both circle landmarks C1 and C2 in all n=43 images from appendix 1, the results of which are summarised in Table 7.1.

Table 7.1
Accuracy results for circle landmark features C1 and C2

Circle landmark C1		Circle landmark C2	
Erroneous landmarks (E_i)	Accuracy ($A_{C1,1}$)	Erroneous landmarks (E_i)	Accuracy ($A_{C2,1}$)
1	0.976	1	0.976

From this data an overall accuracy of 97% is obtained. This is a good result showing that the preprocessing used to filter textural clutter and the application of the smooth kernel Hough transform produces a good localisation of the circle landmarks from all of the species tested.

7.4. Localisation of Landmarks on the Wasp Head Sub-structures

This section discusses techniques for localising the landmark points that reside on the wasp head sub-structures shown in appendix 3. A simple solution would be to use the smooth kernel Hough technique described in chapter 5. Instead of the reference point on each object being the centroid, it could be set to each of the landmarks. However, the maximum which corresponds to the object's reference point can be offset from its true location in the image due to object deformation. Although this is acceptable for object localisation, since we only want to localise the landmark neighbourhood in the image, the resulting landmarks would be displaced giving rise to poor localisation accuracy. It is important that we minimise this problem. Leavers [68] and Stephens [126] discuss the equivalence of the smooth kernel Hough transform with the maximum-likelihood statistical function. Therefore we can calculate, for each dominant curve point within a given neighbourhood, the likelihood of corresponding to the respective landmark point.

7.4.1. Using the Local Shape Structure and Edge Data to Locate each Landmark

This section describes a technique which calculates the likelihood for each landmark candidate from two parameters. The first parameter represents the match between the input data and the sub-structure model. We do not need to match the whole sub-structure to the input data, but only the structure immediately around the landmark point. The second parameter represents the local edge response in the wavelet transform. This is important because the landmark points lie on the boundaries of the wasp head sub-structures. Once the likelihood of each candidate has been calculated for a given landmark point, the candidate with the largest output will be selected as the final landmark.

7.4.1.1. Experimental Design

Using the training set {e100,e200,e201,e202,e204,e405}, mean shapes describing each of the wasp head sub-structures are created. As noted above, only a sub-section of each stored model needs to be used for matching against the input data. If this is too small however, greater ambiguity exists since smaller structures are more likely to reoccur in the input image. For this experiment, the region $[\theta_{li}-\pi/2,\theta_{li}+\pi/2]$ is extracted, where θ_{li} is the orientation of landmark i with respect to the object centroid. The likelihood function described above is then calculated for all candidate points. To test the accuracy of this approach, the landmarks on all of the wasp head sub-structures are localised. The test images used for this experiment are e200, e205, e300 and e400.

7.4.1.2. Algorithm Design

For each candidate point P_i in the neighbourhood of landmark *l* on object *o*, the likelihood of corresponding to the actual landmark point is defined as

$$L_{oli} = gk1D\ (\ 1 - r_i\ ,\ 0.275\ )\ .\ gk1D\ (1- W_{\sigma xy}\ ,\ 0.4\ ) \tag{50}$$

where $gk1D(x,\sigma)$ is a basic gaussian function defined as

$$gk1D\ (x,\sigma) = e^{-x^2 / \sigma^2} \tag{51}$$

r_i represents the binding, or matching of the local shape model to the input data and $W_{\sigma xy}$ is the local edge response for point P_i. The binding parameter r_i is defined as

$$r_i = \frac{M_1 + M_2}{2} \tag{52}$$

where M_1 and M_2 represent the matching of object sub-sections $[\theta_{li}-\pi/2,\theta_{li}]$ and $[\theta_{li},\theta_{li}+\pi/2]$ respectively and are defined as

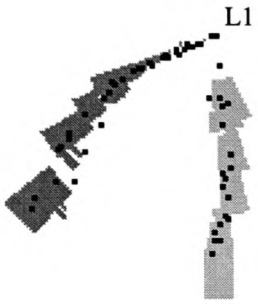
$$M_k = \frac{1}{c_k} \sum_{j=1}^{j \le n} b\ (S_j\ ,\ P_{xy}) \tag{53}$$

where c_k is the number of stored points in object sub-section *k* which are bound to the input points and $b(S_j,P_{xy})$ represents the binding error of stored point S_j to input point P_{xy}.

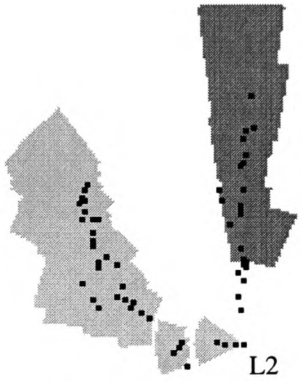
7.4.1.3. Results

Figure 7.6 shows examples of the mean shapes created for the eye and the bottom of the head sub-structures. The shaded regions show the deformation range of each point with respect to the landmark point. This is defined by the training instances used to build the object model. The landmarks found using the above criteria are shown in figure 7.7. These results show poor landmark localisation, with only the top of the eye landmark exhibiting any accuracy and consistency. Using the accuracy criteria described above, a 25% localisation accuracy is obtained. The spatial relationships between each point are not maintained, with landmarks

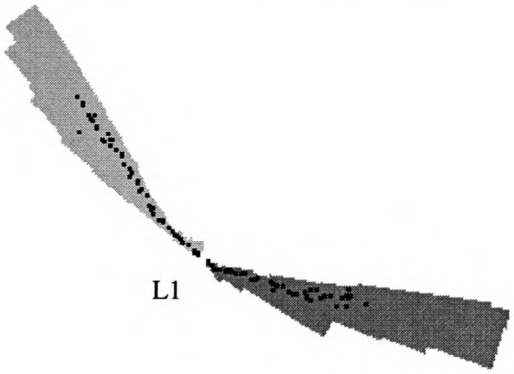
being located arbitrarily along the boundary of each sub-structure. The reason that such offsets occur is due to the problem of clutter in the input dominant curve point map. The function which binds the stored model to the input data, $b(S_j,P_{xy})$, binds to dominant curve points *within* the wasp head model as well as on the boundaries which define each sub-structure. The presence of textural clutter, especially that which follows the contour of the head, gives rise to structural ambiguity, where the local edge response and shape data is insufficient to discriminate between true and false landmark locations. Figure 7.8 shows how textural clutter in images e200 and e400 can give rise to structural ambiguity.



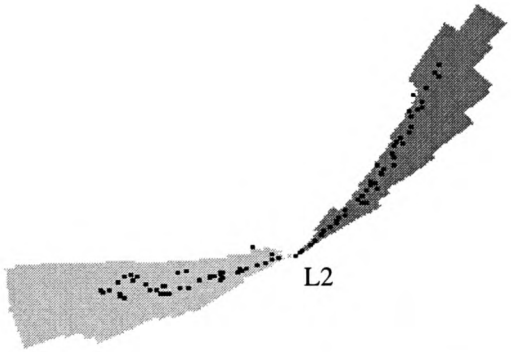
Eye object - shape model around landmark 1



Eye object - shape model around landmark 2



Bottom of head object - shape model around landmark 1



Bottom of head object - shape model around landmark 2

Fig 7.6. Examples of the mean shapes for the "eye" and "bottom of head" sub-structures.

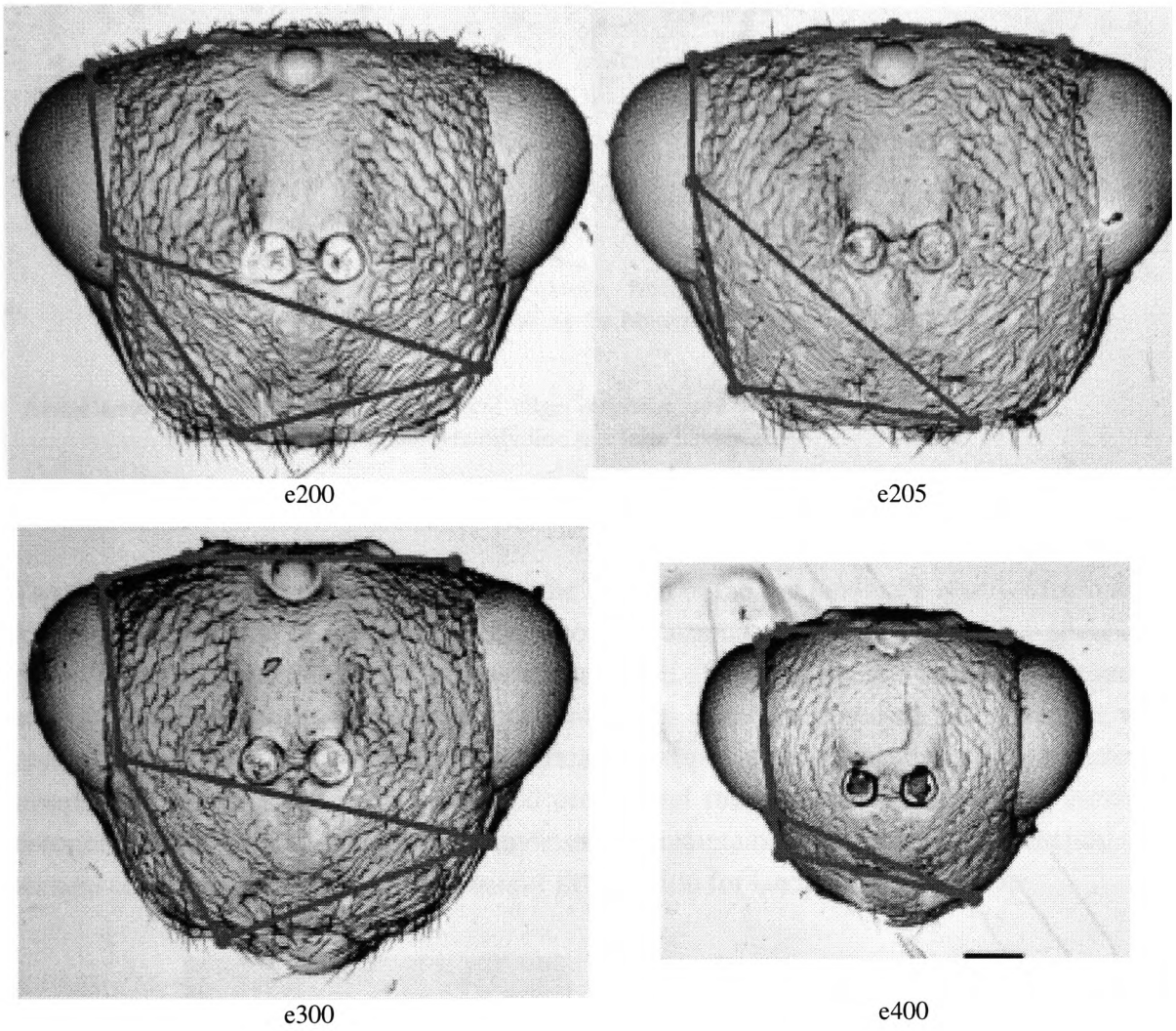


Fig 7.7. Landmark localisation results using the local shape and edge response criteria.

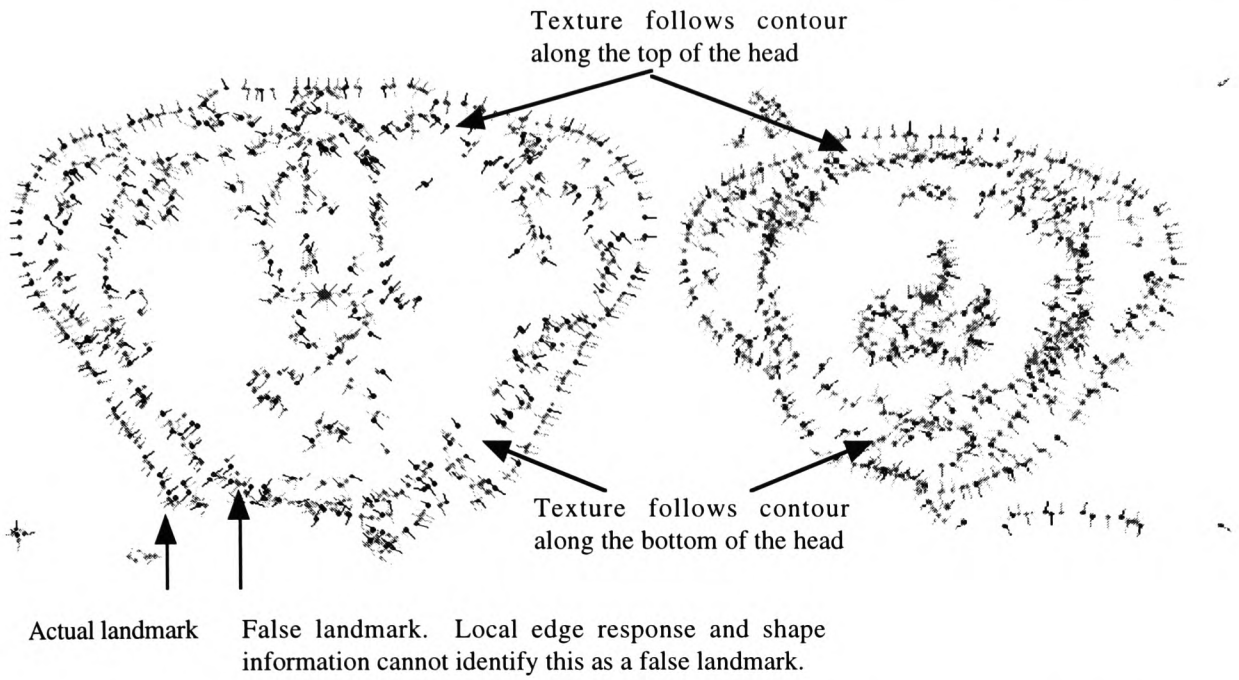
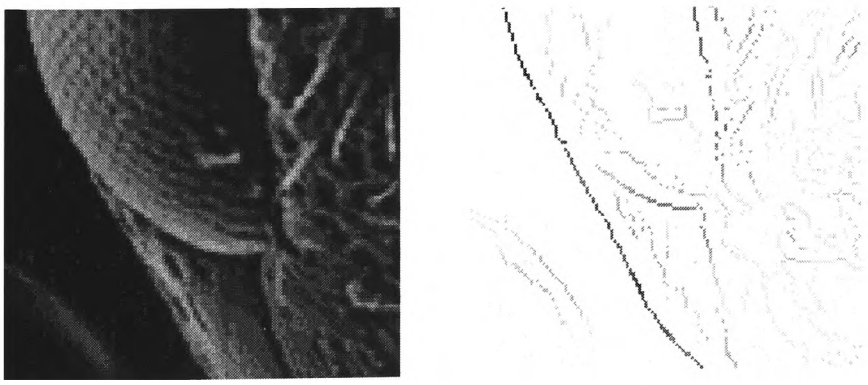


Fig 7.8. Textural clutter that follows the contour of the head can give rise to structural ambiguity.

The results for experiment e200 show that the bottom of the eye landmark is offset from its true location. This is due to the erosion of the contour around that point. Figure 7.9 shows how the contour begins to erode in the wavelet transform. Forming the intermediate contour map attempts to minimise this erosion, but idealised results cannot be obtained. Retaining the whole contour would result in retaining significant textural clutter because such features are inseparable in scale space. This would compound the problems of structural ambiguity described above. This highlights the importance of maintaining a balance between eliminating textural clutter and retaining salient contour information for landmark localisation.



(a) (b)

Fig 7.9. (a) Bottom of the eye for test image e100. (b) Edge map showing how the contour begins to erode around the landmark point. This is an unavoidable effect of gaussian based scale space representations.

Finally, since the structural matching is done with the mean shape, it is important no bias exists. This would result from a large number of similar training instances and the resulting binding to new input data would be inflexible, creating a lower response at points whose local structure did not match the stored object model.

From these results it is clear that local shape and edge data are insufficient to accurately localise the landmark points of interest. To improve these results, three important problems need to be addressed. First, the spatial relationships between landmarks need to be retained. Spatial information from the wasp head model could be used to improve accuracy. Second, textural clutter needs to be minimised further. The above experiments were carried out using dominant curve point maps created with the ROI curve defined in (22). In order to further eliminate textural clutter, the ROI at larger scales of analysis must be larger, as discussed in chapter 4. However, a balance must be sought so salient contour information is not filtered out. Finally, it is important no structural bias exists in the stored object model. In the previous chapter it was shown that training instances which encode the deformation extrema of an object are sufficient. However, it is important that the training instances are evenly distributed over the deformation range. This will allow the above binding procedure to be more robust in the presence of unseen deformation.

7.4.2. Geometric Invariants

In order to improve upon the above landmark localisation results, this section considers the spatial relationships between each of the landmark points. Species are discriminated between based on spatial features that change. However, we want to find spatial features that do not change between each species. From the wasp head models illustrated in figures 5.23 and 6.6(a), it is clear that the spatial relationships between many of the landmark features are not stable. Those features that are stable include the orientation between each of the landmark point pairs on the eye, top of head and bottom of head sub-structures. The orientation of the bottom of the eye landmark with respect to the centre of gravity for the wasp head is also a stable feature. However, the radial distances between landmark points are not stable because the size of the wasp head can change significantly as a result of natural deformation. Figure 7.10 illustrates the above spatial features that can be used to improve landmark accuracy.

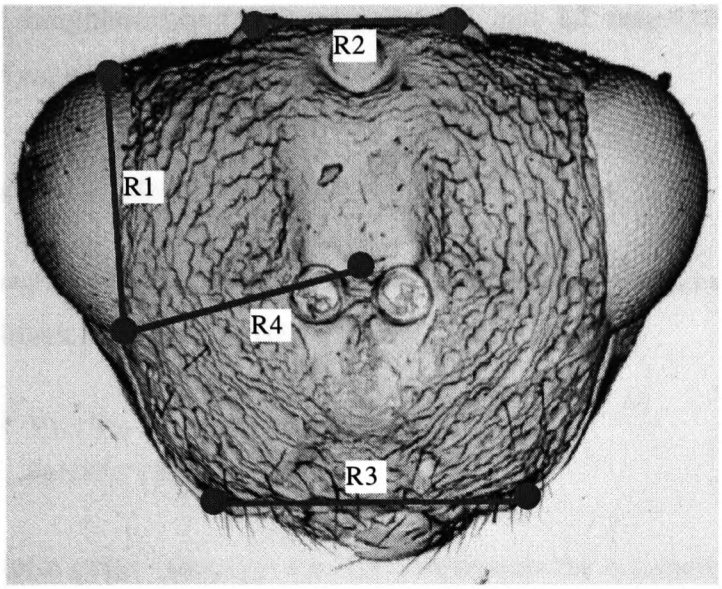


Fig 7.10. Spatial relationships that remain stable between each species. R1, R2 and R3 represent the orientation between the landmark pairs on the eye, top of head and bottom of head respectively. R4 represents the orientation of the bottom of the eye landmark with respect to the centre of gravity of the wasp head.

These features will be incorporated into the likelihood function defined in (50). Symmetry is also an important feature. This will be incorporated into the likelihood function for the landmarks on the top and bottom of the wasp head. Symmetry is not relevant for the eye sub-structure landmarks. Both the training and test image sets used in the previous experiment will be used here to determine if the use of spatial information between landmark points can improve upon the poor localisation accuracy shown in the previous results.

7.4.2.1. Algorithm Design

For the landmark point pairs at the top and bottom of the head, the likelihood function defined in (50) is calculated for both points (L1 and L2) in a given pair. Instead of selecting a single candidate to represent each landmark, two vectors are created, V_1 and V_2, which store the candidate point likelihoods for landmarks L1 and L2 respectively. The likelihood of candidate point i for landmark L1 and candidate point j for landmark L2 are defined as

$$V_{1i} = \text{gk1D}\,(\,1 - r_i\,,\,0.275\,)\,.\,\text{gk1D}\,(1 - W_{\sigma i}\,,\,0.4\,) \tag{54}$$

$$V_{2j} = \text{gk1D}\,(\,1 - r_j\,,\,0.275\,)\,.\,\text{gk1D}\,(1 - W_{\sigma j}\,,\,0.4\,) \tag{55}$$

Let Θ define a matrix of order (n1xn2) where n1 and n2 represent the number of candidate

input points in the neighbourhood of landmarks L1 and L2 respectively. The spatial relationship between each point Θ_{ij} is defined as

$$\Theta_{ij} = gk1D\ (\ \theta_{ij}\ ,\ 0.4\ ) \tag{56}$$

where θ_{ij} is the orientation of landmark L1 candidate *i* with respect to landmark L2 candidate *j*. For all possible landmark point pairs (i,j) the likelihood function

$$\begin{aligned} L_{olij} = \ & V_{1i}\ .\ V_{2j}\ .\ gk1D\ (1 - \theta_{ij}\ ,\ 0.2)\ . \\ & gk1D\ (\ ((\ |\ P_ix - Cx\ |\ ) - (\ |\ P_jx - Cx\ |\ ))\ ,\ 0.2) \end{aligned} \tag{57}$$

is maximised. The term $(\ |\ P_ix - Cx\ |\ ) - (\ |\ P_jx - Cx\ |\ )$ represents the symmetry of each landmark point with respect to the centre of gravity for the wasp head. P_ix and P_jx represent the *x* positions of landmark candidates *i* and *j* respectively while C_x represents the *x* position of the centre of gravity. The σ parameters to the gaussian functions are smaller for the new spatial features, giving them more emphasis in the final result. Given (i,j) that maximises the above function, the landmark point pairs on the top and bottom of the wasp head sub-structures can thus be defined as

$$L1 = i \qquad L2 = j \tag{58}$$

For each candidate point in the neighbourhood of the bottom of the eye landmark (L2), the likelihood function is defined as

$$\begin{aligned} L_{12} = \ & gk1D\ (\ 1 - r_i\ ,\ 0.275\ )\ .\ gk1D\ (\ 1 - W_{\sigma xy}\ ,\ 0.4\ )\ . \\ & gk1D\ (\ |\ \theta_{cl} - \theta_{ci}\ |\ ,\ 0.25\ )\ .\ gk1D\ (\ |\ \theta_{L12} - \theta_{L1i}\ |\ ,\ 0.25\ ) \end{aligned} \tag{59}$$

where θ_{cl} and θ_{ci} represent the mean orientation of the landmark in the stored eye model and the orientation of candidate input point P_i with respect to the centre of gravity respectively. Since the landmark at the top of the eye sub-structure (L1) is stable, it can be used to localise the second landmark. θ_{L12} and θ_{L1i} represent the mean orientation of landmark L2 in the stored eye model and the orientation of candidate input point P_i with respect to the top of the eye landmark.

7.4.2.2. Results

Figure 7.11 shows the landmark localisation results for the new spatial relationship criteria applied to the test images e200, e205, e300 and e400.

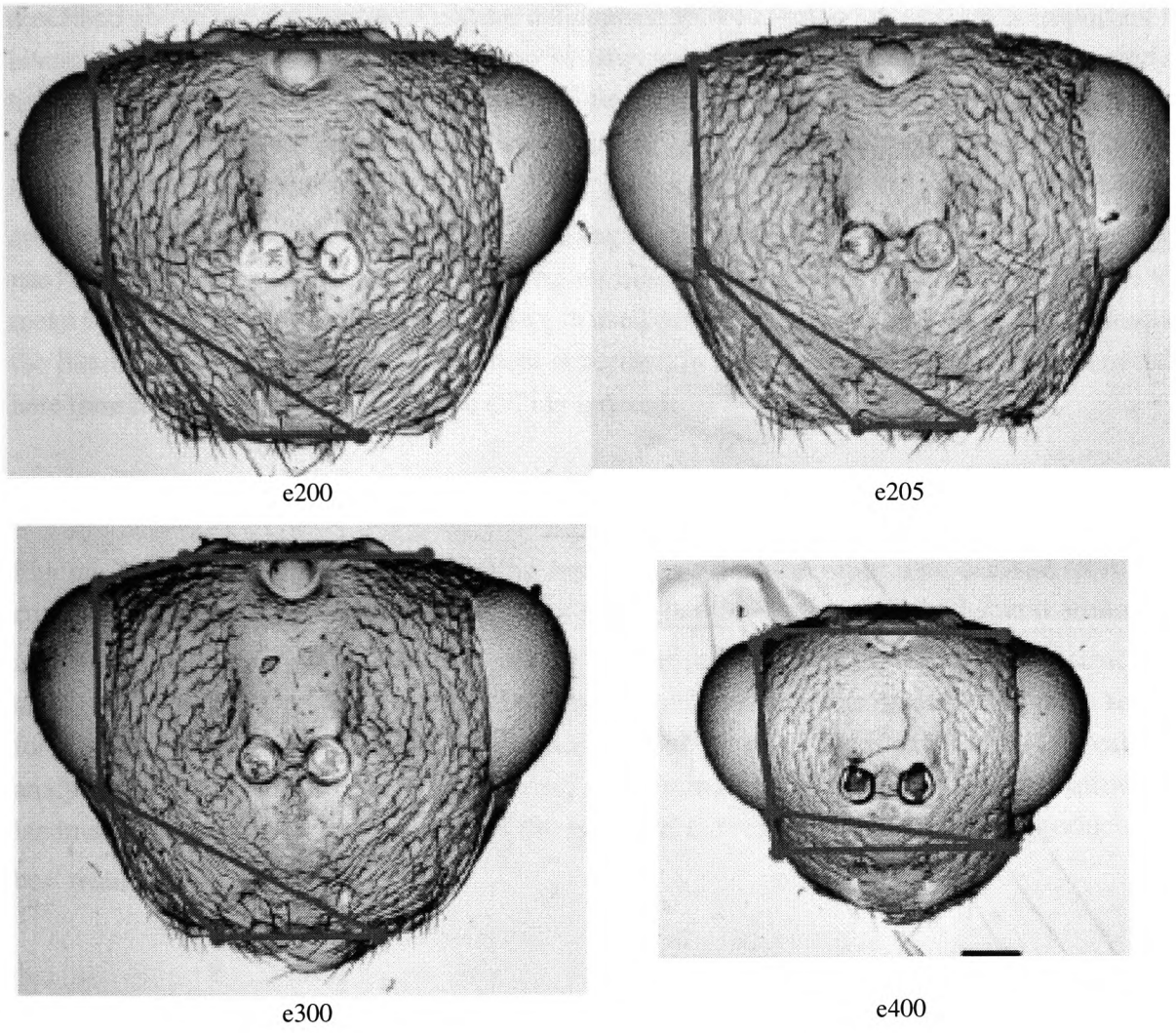


Fig 7.11. Landmark results which take into account the spatial relationships between each point.

These images show an improvement over the original localisation results. Here a localisation accuracy of 31% is obtained. This is not acceptable since the landmarks are still being offset from their true locations. This shows that false landmarks within each neighbourhood can satisfy both the spatial relationship and symmetry criteria. Although these criteria ensure landmarks retain their relative spatial positions, clutter in the input dominant curve point map still gives rise to false landmarks which remain spatially valid.

7.4.3. Eliminating Correlated Noise

Since textural clutter gives rise to structural ambiguity, as shown in figure 7.8, it is important this is filtered out so the ambiguity is eliminated. However, it is important salient contour information is retained, especially around the bottom of the eye sub-structure. The experiments described above use the contour / texture delineation ROI as defined in (22). It is important to investigate how localisation accuracy can be improved when the ROI is increased in size at larger scales of analysis. This was shown to filter a significant amount of textural clutter which exists at larger scales of analysis in the wasp head images. To investigate this, two ROI curves ε_1 and ε_2 are tested, where $\varepsilon_1=\{16,32,48,64\}$ and $\varepsilon_2=\{32,64,128,128\}$. These are used to create the intermediate contour map and resulting dominant curve point map for the test images used in the above experiments. The training set used in the previous experiments to derive the mean shape models for structural matching is used here also. The algorithm used to calculate the likelihood of each landmark candidate described in section 7.4.2.1 is used here also, only here the symmetry criteria in equation (57) is ignored.

7.4.3.1. Results

Figures 7.12 and 7.13 show the resulting landmarks given the input data derived from ROI curves ε_1 and ε_2 respectively. These results show that the accuracy of the located landmarks improves significantly as the ROI increases in size at larger scales. Clearly, the structural ambiguity which caused the problems described above has been minimised. The main reason for this is that a significant amount of detail on each of the wasp heads exists at larger scales of analysis. Hence a larger ROI is necessary to eliminate irrelevant features and allow the landmarks to be accurately located. Of the two ROI curves used here, curve ε_2 produces the best results.

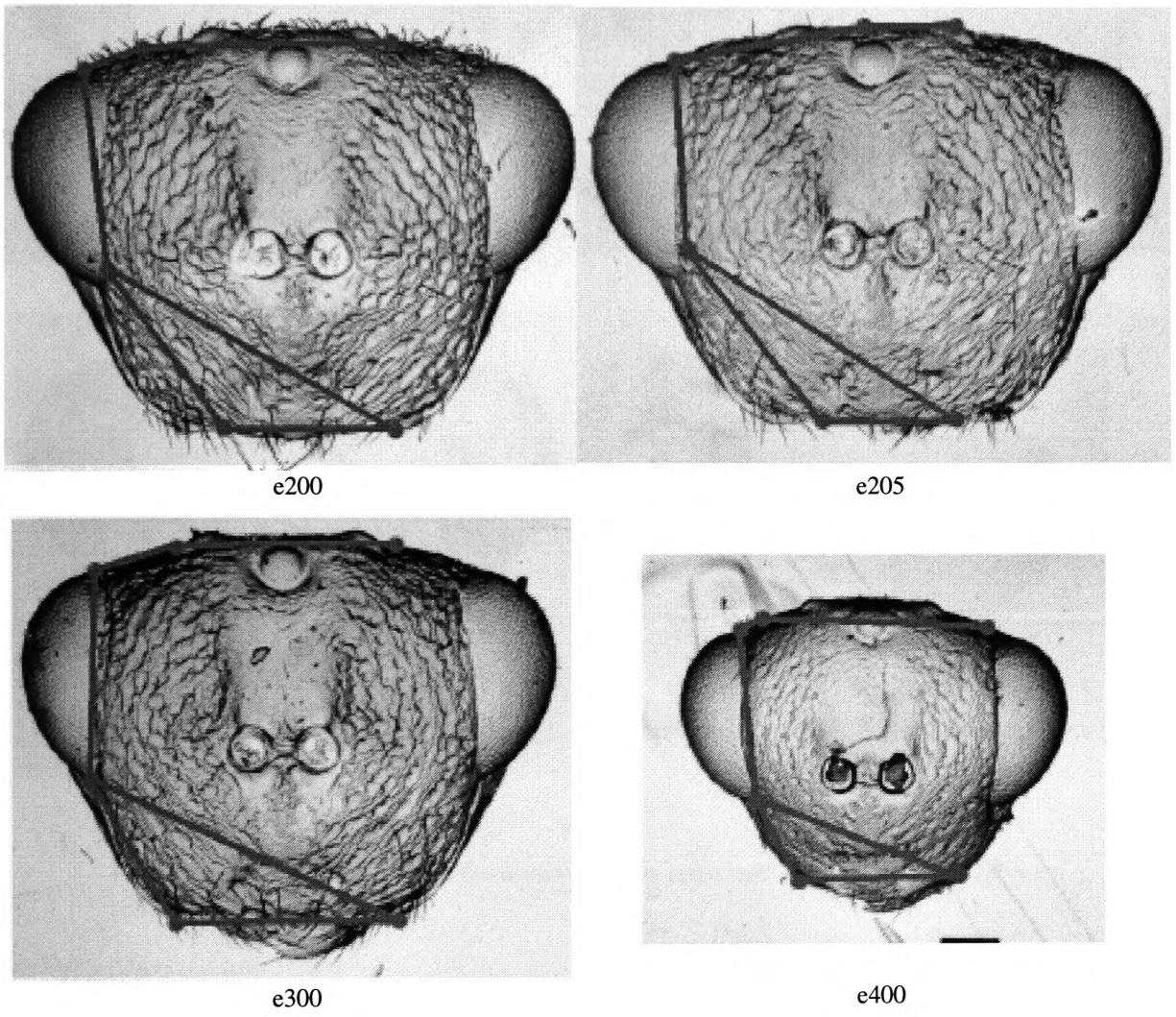


Fig 7.12. Landmark results from the dominant curve point map derived from ROI curve ε_1.

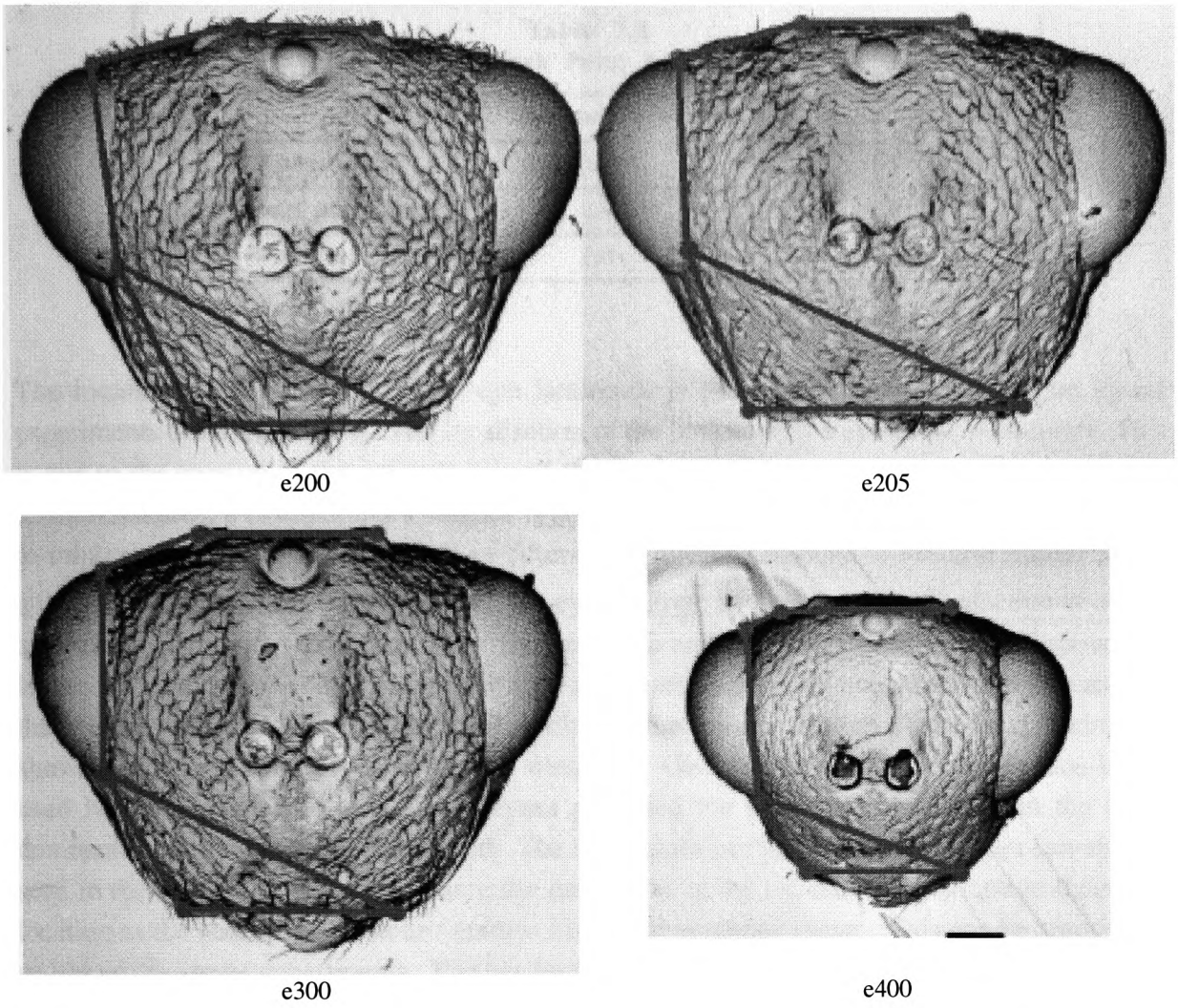


Fig 7.13. Landmark results from the dominant curve point map derived from ROI curve ε_2.

The criteria selected for determining the accuracy of the resulting landmark locations is the same as that used for the circle landmarks, and can be quantified using equation (49). Appendix 8 shows the final set of landmarks created using ROI curve ε_2 while table 7.2 summarises the landmark accuracy.

Table 7.2
Landmark Point Accuracy

	Landmark L1 Accuracy	Landmark L2 Accuracy
Eye substructure	0.976	0.83
Top of head substructure	0.93	0.93
Bottom of head substructure	1.0	1.0

The localisation for the top of the eye landmark is good, as was the case in the initial experiments described above. The localisation of the bottom of the eye is not as accurate. This is due to the erosion of the contour around the landmark point. When filtering contour and textural features, it is important a balance is achieved. The algorithm described above attempts to minimise this offset. ROI curve ε_2 filters a significant amount of textural clutter while attempting to minimise the erosion of the eye contour and subsequent displacement of the associated landmark. The results for images e502 and e503 in appendix 8 also show that landmarks can also be found when the object of interest is partially occluded, as is the case for the eye sub-structure. The localisation results for the top and bottom of the head landmarks shows a high degree of accuracy can be obtained. Given that the symmetry criterion is not used here, it shows that is is not relevant provided the structural ambiguity in the input dominant curve point map is minimised. The importance of filtering textural clutter can also be seen in the case of image e400, where the landmarks at the top of the head tend to their true location as the size of the contour / texture filter ROI becomes larger. This can be seen in the results of the above experiments. Further filtering would improve the results so the landmarks could be correctly located. However, this would likely erode other contour features, causing localisation problems at other landmark points. The results in appendix 8 also show landmarks can be located when irrelevant clutter is present around the location of the landmark. The fidelity of the intermediate dominant curve point map is also as important as the reduction in its textural content. The top of the head sub-structure in image e505 obscures the landmark point with a hair-like texture. This is represented as a single contour in the dominant curve point map, and as a result, landmark L1 on the top of the head sub-structure is offset slightly from its true location.

7.4.4. Conclusion

The above results show that the spatial relationships between each landmark point, the local structure and edge response are important criteria for locating taxonomic landmark points. However, the properties of the wasp head image means that significant textural clutter resides at larger scales of analysis and is inseparable from salient contour information. It has been

shown that the filtering of textural clutter is crucial for the accurate localisation of landmark points, eliminating the structural ambiguity which gives rise to false landmark candidate points. Since large scale clutter and important edges are inseparable in scale space, a balance must be achieved in eliminating textural clutter and retaining important contour features. This is done with ROI curve ε_2, where the erosion of the bottom of the eye and subsequent displacement of the associated landmark point is minimised. Finally, taking into account the data in tables 7.1 and 7.2, an overall landmark accuracy of 95% is obtained, with the landmarks that constitute the 5% error showing only a small offset from their true location. These findings represent the main contribution to knowledge for this work and serve as the basis for automating the identification of each wasp species.

Chapter 8

Species Classification

8.1. Introduction

The final process of any image analysis system uses the high level data representations extracted from an input image to perform a given task. This may be the control of another system, robot navigation for example [73], or the classification of objects in the image. This work focuses on the latter, using the taxonomic landmark points extracted in chapter 7 (and shown in appendix 8) to classify each species of parasitic wasp head. There are many statistical techniques that can be applied to solve this problem [11]. Artificial neural networks, themselves a statistical mechanism, have also been widely used for problems of biological and medical image classification [62,114,133]. The inherent advantages of neural networks are their robustness to variable and noisy input data. Given that, once trained, neural networks are robust and fast in their execution, they will be used here to perform the final classification of the wasp head images. Just as there are many statistical techniques, so there are many neural network architectures and learning methodologies. Here one of the simplest forms of network architecture is used. This is the back propagation neural network [32]. This uses a supervised learning technique, where the weights are adjusted according to the error between the actual and expected output of the network for a given input pattern.

8.2. Experimental Design

Given the localised landmark points shown in appendix 8, the taxonomic features illustrated in appendix 2 are calculated for each wasp head image. These 28 features represent the complete set of taxonomic features used by experts for species classification and form the input vector to the neural network. The network will have 5 output nodes, each one corresponding to a particular species. To determine the classification accuracy, or recognition rate of the network, the set of experiments summarised in table 8.1 will use a different number of hidden nodes, training epochs and wasp head instances in each training set. The test set for each experiment comprises the taxonomic feature vectors for the complete set of images shown in appendix 1. The implementation of the neural network is taken from Fausett [32]. The transfer function of each node, which maps the input sum to a normalised output range, can be binary [0,1] or bipolar [-1,+1]. The latter has been shown by Fausett [32] to be more effective during training, allowing for better convergence and a smaller network error rate. Therefore, the transfer function of the back propagation network implemented here will be a bipolar sigmoidal curve, an example of which is shown in figure 8.1. The values of each input vector are normalised to avoid problems of saturation. Here the transfer function tends asymptotically to its extrema (-1 and +1 in this case) and the resulting first derivative function used to modify the network weights during training will be close to zero. This is also shown in figure 8.1. Hence the modification of the network weights will be negligible, resulting in poor convergence and a

larger network error [32].

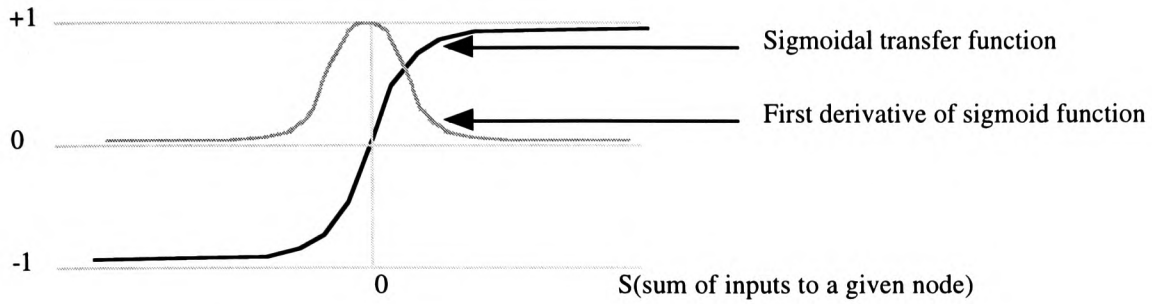


Fig 8.1. A bipolar transfer function improves the classification performance of the network. For un-normalised input data the total input to each node (S) will be large, resulting in a small derivative value. The weight updates during training will thus be negligible and the network's error rate will be high.

Table 8.1
Species Classification Experiment Parameters

Experiment	Hidden nodes	Epochs	Training set
1	16	4000	{e100,e101,e105,e107,e200,e202,e204,e300,e302,e305, e400,e405,e406,e505,e507}
2	30	10 000	{e100,e101,e105,e107,e200,e202,e204,e300,e302,e305, e400,e405,e406,e505,e507}
3	30	10 000	{e100,e101,e105,e107,e200,e201,e202,e204,e300,e301, e302,e305,e400,e401,e405,e406,e500,e501,e503,e507}
4	40	10 000	{e100,e101,e105,e107,e200,e201,e202,e204,e300,e301, e302,e305,e400,e401,e405,e406,e500,e501,e503,e507}
5	40	14 000	{e100,e101,e105,e107,e200,e201,e202,e204,e300,e301, e302,e305,e400,e401,e405,e406,e500,e501,e503,e507}

8.3. Results

The classification results for each experiment are shown in tables 1 to 5 in appendix 9. Each row corresponds to a particular test image while the columns represent different species. The numbers in each row represent the response of the 5 output nodes from the network. Since a bipolar transfer function is used, -1 represents a poor match while values closer to +1 represent a good match. The output response for the nodes which correspond to the species each test image belongs to are shown in a bold typeface. The results for experiment 1 show a recognition rate of 76%. A significant number of errors occur for the images in species 2, species 3 and species 5. This is due to the similarity between each of the species which results

in the overlap of the species' clusters in the taxonomic feature space. Here images e201, e203 and e206 are misclassified as species 3 while image e205 is misclassified as species 5. Images e301 and e306 show a large output for species 5 while images e500,e504 and e501,e506 are misclassified as species 2 and species 3 respectively. Species 1 and 4 exhibit some degree of overlap, with image e403 being misclassified as species 1.

Ideally, each wasp head lies within an easily identifiable cluster of similar instances that comprise a given species within the taxonomic feature space. This is not the case so it is important to appropriately delineate the feature space. The hidden nodes of the neural network act as feature detectors, each defining a hypercurve which partitions the multi-dimensional taxonomic feature space. The greater the number of hidden nodes, the better the delineation of species clusters. However, too many hidden nodes limits the networks ability to generalise unseen input data. Experiment 1 uses only 16 hidden nodes. For experiment 2, this is increased to 30 and the number of training epochs is increased to 10,000. The results, summarised in table 2 show no significant improvement, with the recognition rate remaining at 76%. The only difference is that the result for image e306 increases significantly. For the node corresponding to species 3, the output is 0.54 where it was recorded as -0.15 in experiment 1. However, the response for the node corresponding to species 5 is higher, with an output of 0.64. This results in the same misclassification as in experiment 1.

Aside from the number of hidden nodes, the number of training instances used is also important. The above two experiments use the same training set. To determine if the results can be improved, a new training set is used for experiment 3, with an extra instance added for species 2,3 and 4, and two instances added for species 5. This is shown in table 8.1. The results for this experiment are summarised in table 3. A recognition rate of 86% is now obtained. This is a significant improvement over the above results, showing that an increase in the number of training instances can result in an increase in the recognition rate of the neural network. Experiment 4 uses the same training set but the number of hidden nodes is increased to 40. Table 4 shows that a recognition rate of 91% is obtained. Again this is a significant improvement and it shows that more hidden nodes are required to delineate the taxonomic feature space given the overlap between species shown above. Discussions with experts reveals they can achieve a recognition rate of around 65%. Clearly, a significant improvement can be obtained using the neural network architecture. However, species 2 still exhibits significant overlap with species 3. From the results of experiment 4, images e203, e205, e206 and e505 are all misclassified as species 3. Analysing the landmarks shows that they are located accurately for these images. Where significant landmark displacement does occur, such as the circle feature C1 in image e501, the image is classified correctly. Although the displacement of those landmarks which are offset from their true location is small, this result shows that the neural network is robust to the slight error in landmark localisation as a result of contour erosion and false maxima in the Hough space. Looking at the images for species 2 and

3 shows they are visually similar. This similarity can also be shown statistically. Table 6 shows the distance between each image in species 2 and species 3 in the taxonomic feature space. The mean distance between species 2 and species 3 is calculated as 0.36. If we look at the distances between instances within species 2 alone, as shown in table 7, a mean distance of 0.34 is obtained. Clearly this is not a significant difference, especially compared to the mean distance between two dissimilar species, such as species 1 and 3, as shown in table 8. Given that the mean distance of image e206 to the images of species 3 in table 6 is 0.27, it is clear that both species 2 and species 3 overlap significantly. A similar overlap between species 5 and species 3 is also shown in table 9. Here the mean distance between species 5 and species 3 is 0.48. However, image e505 has a mean distance of 0.35 from the instances that comprise species 3. Such an overlap results in the misclassification of this wasp head instance. To overcome these problems, the number of training epochs is increased to 14,000 in experiment 5. However, the results in table 5 show a recognition rate of 90%, where image e506 is now misclassified as species 2. This result highlights the problem of over-training the network, where the training data is presented too many times. The resulting network becomes less robust to variable input data. Handling variable data is important for this work, so the results from experiment 4 show that around 10,000 epochs are suitable for the data used here.

8.4. Conclusion

A back propagation neural network has been implemented to classify the wasp head images from the complete set of 28 taxonomic features derived from the landmark data created using the statistical techniques described in chapter 7. A recognition rate of 91% has been obtained given a sufficient number of training instances, hidden nodes and training epochs. Discussions with experts in the taxonomy of parasitic wasps reveals that they can obtain a recognition rate of around 65%. Clearly, the recognition rate of the network is significantly higher. The number of hidden nodes is important to delineate the taxonomic feature space given the significant amount of overlap between each species. However, it is important that a balance is achieved because too many hidden nodes results in the poor generalisation capability of the network and too few hidden nodes results in a poor recognition rate as shown in the above results. Misclassifications were shown to occur for images where the overlap was greatest. The network is robust to the landmark localisation error introduced as a result of contour erosion and false maxima in the Hough space. As such the neural network described here is deemed a suitable mechanism for high level processing in the species classification system described throughout this work.

Chapter 9

Discussion and Conclusions

9.1. Discussion

This work has studied the problem of automating the classification of parasitic wasps of the order *Hymenoptera* using taxonomic landmark points which are automatically extracted from SEM images of the anterior view of the wasp heads. Various image feature extraction, object localisation and statistical matching techniques have been investigated in order to automate the landmark localisation and species classification processes. This section discusses the results obtained and considers whether the goals set out in chapter 1 have been achieved.

9.1.1. Image Feature Extraction

Initial work focused on techniques which are described throughout the literature, looking at how deformable contour models, facial feature extraction, texture analysis and multiscale edge representations can be applied to the problem of taxonomic landmark localisation. Deformable contour models were shown to perform poorly under conditions of severe image clutter and as such the resulting wasp head sub-structure and landmark localisation accuracy would be sub-optimal. Given the significant amount of textural clutter around the sub-structures of interest on the wasp head, such techniques would fail. Indeed, the use of raw image data to extract any localised features is likely to fail given the significant textural variance around the landmark points of interest and the natural deformation between each wasp head instance. Given the amount of textural detail in the images, texture analysis techniques which cover both the segmentation and classification of textural regions have been studied in detail. Although intuitively such techniques would be applied to the wasp head images, the problems of poor edge localisation and the introduction of correlated noise through the identification of false regions, which are recognised throughout the literature and illustrated in figure 3.1, make them unsuitable for landmark localisation. Firstly, it is vital that any intermediate representation retains accurate edge locations, since landmarks reside on the boundaries of salient structures on the wasp head. Poor edge localisation would effect the accuracy of the landmark locations. Secondly, texture analysis techniques introduce irrelevant clutter. The results described in chapter 7 show that such clutter should be minimised to overcome the problems of false landmark points. Also, the SEM images under consideration present two types of taxonomic landmark, points on object boundaries and whole object structures. The latter correspond to the circular structures (C1 and C2) located in the top and mid regions of the wasp head. Considering these as texture reduces to correlation [73]. Although this is robust to the global translation and rotation transformations [105], it is sensitive to local object deformations which are prevalent in the image data under consideration. Using multiscale edge representations derived from gaussian based operators allows textural clutter to be filtered out. The causality property states that no new detail can emerge in the resulting scale space. This overcomes the

problems of introducing correlated noise inherent in texture analysis methods. However, the problem of poor edge localisation still remains due to the blurring required to filter out smaller scale features. Edge events are displaced from their true location and can split and merge with other events at bifurcation points. However, this problem is easily overcome because each edge event can be tracked across scale space to its true location at the smallest scale of analysis. The results highlighted in figures 3.3 and 3.5 show that edges can be moved back to their true location and the inherent deformation of each contour can be eliminated. Such information is not available in the edge map derived from texture analysis methods, so the edge displacement cannot be resolved so easily. The Mallat wavelet transform has been applied here because it is a first derivative operator. As such, edge events at each scale of analysis can be used to differentiate between salient image features and irrelevant clutter caused by image noise. Hence the Mallat wavelet transform can be used to achieve the first goal set out in chapter 1, allowing the edge features that define each landmark to be accurately localised.

9.1.2. Delineating Contour and Textural Features

The creation of the multiscale representation results in an expansion of the data that is available, with 4 separate edge map images being created over the dyadic scale range $[\sigma=2^0,\sigma=2^3]$. To overcome this problem an appropriate scale can be selected for subsequent processing [81]. However, the multiscale representation created here presents two important problems. First, although larger scales of analysis eliminate textural clutter, salient edges that define important sub-structures on the wasp head can erode, as shown in figure 7.9. Second, if we use smaller scales of analysis, where contour erosion does not occur or is less prevalent, textural clutter is reintroduced. To overcome these problems it has been acknowledged that salient contour features must be separated from irrelevant textural clutter. The various techniques based on the localised variation between cellular textures [25] are unsuitable given the natural variation in the texture being analysed. Recent work by Ganesan and Bhattacharyya [38] attempts to find edges in textured images, but their results show poor edge localisation. To overcome this problem, a novel technique is described which recombines the edge map images created with the Mallat wavelet transform to produce a single contour map, maximising the presence of salient contour features that define the wasp head sub-structures (and overcoming the problem of erosion) while minimising textural clutter. Given the correlation properties of the wavelet transform as described in chapter 3, the emergence of edge events in a coarse to fine scale traversal, the properties of the wavelet transform modulus maxima and the local activity along the gradient of each wavelet response are deemed important characteristics in distinguishing contour and textural features. Figure 4.2 illustrates how these criteria can be used to identify contour features. Edges between smooth regions are defined by isolated contour events with little or no surrounding activity. Textured regions can be bounded by texture of different spatial frequency and orientation [23] or by smooth regions. The latter results in the decay of

local activity around contour points while the former presents a class of boundary which cannot easily be identified from the wavelet transform itself, and as such requires a two-phase approach where texture region segmentation is applied to the identified textural regions.

Initial experiments were applied to SEM images of the anterior view of the head of a Fruitfly *(Drosophila melanogaster)*, shown in figure 1.1(a) and a synthetic image containing both contour and textural features shown in figure 4.7(a). The interaction between neighbouring maxima at each scale was calculated within the ROI curve defined in (22). Figure 4.6 and 4.7(b) show the final contour maps for each of these images respectively. Significant textural clutter has been eliminated and the data available for subsequent processing is reduced to a single contour image. Figure 4.5 shows that salient contour features can be identified, even at the smallest scale of analysis where a significant amount of textural clutter resides. We see that features on the perimeter of the textured region in figure 4.7(a) are identified as contours. This results from the decay in interaction around the border of textured regions which neighbour continuous regions. However, figure 4.7(b) shows that salient contour features have also been filtered out. This problem arises from the interaction of two edge events in close proximity, making it difficult to distinguish between structures of sufficiently small spatial extent and regions of significant textural activity. For the wasp head data this is not a significant problem and only occurs for some instances of the circular landmark feature C1, as shown in figure 7.3. The same experiment was also carried out for a set of 6 wasp head images from appendix 1. The results illustrated in figure 4.8 show that significant textural clutter remains. By configuring the ROI curve so that it is sufficient large at the high end of the scale range improves the results significantly, as shown in figure 4.9. This is because a significant amount of textural clutter resides at larger scales in the wasp head images. Better results were obtained for the Fruitfly head and synthetic image using the original ROI curve because they do not contain significant textural detail at larger scales of analysis. As such the exponential ROI curve defined in (22) was sufficient. However, the results in figure 4.9 show that some textural clutter remains. Increasing the ROI or scale range would filter this clutter, but important contour features that define the wasp head sub-structures would also begin to erode. This is because these features are inseparable in scale space. It is therefore important to achieve a balance when configuring the ROI curve, so that both textural clutter and contour erosion are minimised.

9.1.3. Contour Model Extraction

A significant drawback of using multiscale edge representations is that contours become discontinuous at larger scales. This is the result of edge displacement inherent in the gaussian smoothing process. Textural filtering also results in the erosion of important contours. As such the final contour map contains many fragmented contour segments. To overcome this

problem the existing segments in the contour map must be reconnected. This can be done using the more complete edge information in the wavelet transform. However, this introduces two fundamental problems. First, extending contour segments using the edge information at the smallest scale of analysis, where all edge data for the image is contained, causes problems because extending contours within cluttered environments is prone to failure [65]. Given the wasp head data available here, significant textural clutter resides at the smallest scale. Secondly, extending contour segments can reintroduce unwanted textural clutter. As such, any contour extension process must be suitably constrained so we do not augment false contours and reintroduce textural features. Figure 5.6 illustrates the above problem, where significant textural clutter is reintroduced into the contour map. Hysteresis thresholding can be used to eliminate contours that are defined by noise [13], but this cannot be applied here because contours that define textural features can retain significant edge point responses that remain above the hysteresis threshold, as shown in figure 5.7. From the contour maps shown in appendix 5, discontinuities between contour segments that define important structures on the wasp head are only one or two pixels in size while reintroduced textural segments are significantly longer. A statistical model of contour lengths is presented which constrains the reintroduction of textural clutter by using the mean (reintroduced) segment length as a threshold. When reintroduced segment lengths are greater than the threshold they are removed from the contour map. Figure 5.10 shows that the reintroduction of textural clutter can be significantly reduced compared to the result in figure 5.6. If the threshold is too low or too high then salient contour features are eliminated or irrelevant clutter is retained, as shown in figures 5.11 and 5.12 respectively. Certain images may be the exception, as is the case for image e105 shown in appendix 5. Here a border around the head resulting from background image noise causes the mean segment length to be much higher. As a result more textural clutter is retained in the final contour map. However, the landmark localisation results described in chapter 7 and shown in appendix 8 are not seriously affected by the extension of textural segments as they do not necessarily introduce structural ambiguity, but are more sensitive to the configuration of the ROI curve used to filter textural clutter.

9.1.4. Wasp Head Shape Representation

Since the taxonomic landmarks of interest reside on the boundaries of wasp head sub-structures, it is important to appropriately model these boundaries for subsequent landmark localisation. Many representations described throughout the literature are based on 1D curve models, using curvature scale space [65] to extract the general curvature properties of a given shape and eliminate boundary noise. The nature of the SEM wasp head images means we are unlikely to form complete (closed) curves in the contour map that describe the wasp head sub-structures. This problem is compounded by the constraints applied to the contour extension process to overcome the problems of erroneous edges and textural clutter being reintroduced

into the contour map. As such, the contour map is likely to contain fragmented curve segments, none of which represent a complete shape. Even if a set of contour segments could be extracted that represent a complete shape, additional part-whole problems become relevant [89]. The curvature of a closed form indicates concave / convex points with the positive and negative sign of the curve function respectively [65]. If the boundary point order used to calculate the curvature function is inverted, then the concavity / convexity is also inverted, as shown in figure 5.14. For isolated contour segments, this ambiguity would be difficult, if not impossible to resolve. Second, in creating the curvature scale space for isolated contour segments, the curve maxima at the ends of the segment, where the smoothing function goes off the edge, would vary significantly, giving rise to inaccurate results [89]. This problem is compounded if the periodicity of open curves is not 2N, where N is the length of the curve. Figure 5.15 shows how no periodicity, 1N and 2N periodic open curves behave over scale space. Zero or 1N periodic curves result in severe shape deformation over the scale range, while 2N periodic curves, which are symmetrical at their end points, retain the curve shape and fidelity of the curvature scale space. Also, as a result of smoothing, curvature scale space distorts the curvature at larger scales, as shown in figure 5.13. Mokhtarian and Mackworth [89] note that this does not detract from the important properties of the representation : Invariance, uniqueness and stability. However, the curvature scale space they derive is based on second derivative zero crossings, not first derivative maxima. Since this contains only position information, it is not sensitive to the maxima variance in open curves that is described above. So although maxima are useful for low level image representations, for intermediate shape / contour representations, zero crossings are more useful. Since we cannot assume that important features will exist or conditions will be met in the input data, image analysis processes and the representations used for pattern matching and object localisation must therefore be robust to variable and noisy input data. As such a curvature scale space representation based on zero crossings would be more robust to the variable properties of the input contour map.

Since we cannot form complete 1D curve representations using the zero crossings at dominant curvature points, we can map these points from local curve space back into the image plane. This provides a distributed representation of salient shape data in the form of a dominant curvature point map. Appendix 6 shows the dominant curve point maps for the wasp head images used for this work. Due to the above problems, it is difficult to make assumptions about which subset of dominant curve points define each object of interest and the connectivity between these points. Therefore a convenient 1D curve representation cannot easily be adopted. Instead, a polar-curvature representation is used. Such a representation can be generalised to represent any arbitrary shape [122]. Deformation can be encoded by finding correspondences between points on each object instance. Since all instances share that same taxonomic landmark points, binding these facilitates the mapping of non-landmark points in the polar coordinate space. Although the representation is generic, the mechanism for binding

shape instances is not. Here only shapes which satisfy the centroidal profile constraint [122] can be bound. However, the wasp head sub-structures do satisfy this constraint, as shown in figure 5.17. The combined object representation created by binding together individual instances is over complete, as shown in figure 5.21. This contains redundancy due to the emergence and disappearance of dominant curvature points as a result of local deformation between object instances. However, this is important since it encodes the behaviour of dominant curvature points over the deformation space which the object training instances define. Using this representation, the second goal set out in chapter 1 has been achieved.

9.1.5. Wasp Head Sub-structure Localisation

In order to localise the taxonomic landmark points, it is important to define a local neighbourhood in which each landmark resides. Segmenting the search space in this way reduces the time required to search for the landmark points and reduces ambiguity [68]. Given the limitations of deformable contour (snake) models, two other techniques have been used to locate the eye, top of head and bottom of head sub-structures. These are elastic graph matching and the Hough transform. Both lend themselves to the dominant curvature point map used to represent the input image data. Elastic graph matching produces good results when little or no deformation is present, as shown in figure 6.1, but poorer results are produced when significant deformation is present. Figure 6.2 shows that the outer part of the eye object is stable and thus produces a good result, but the inner part of the eye produces a poor result. This is because the unseen deformation in the test images deviates significantly from the deformation encoded with the training instances. This problem can be overcome by ensuring the deformation extrema of each point is contained in the training set, or by using the deformation properties of each point to extrapolate their deformation range. Computationally the elastic graph matching algorithm is expensive. This is because assumptions cannot be made about the location of object in the image, so all input dominant curve points must be analysed. Since we are only looking for landmark points, it is also performing unnecessary processing by localising all of the points on a given object.

The Hough transform is widely used for locating structures in images, and the generalised Hough transform [7] is utilised here since the objects of interest on the wasp head cannot be represented analytically. The dominant curve point representation also lends itself to the generalised Hough transform, as Huttenlocher and Grimson [43] note that greater success can be obtained if vertex data is used instead of continuous edge points. Two variations of the generalised Hough transform are used. First, the conventional approach, as proposed by Ballard [7], uses a pixel wide "top hat" function to modify the accumulator array. Secondly, the smooth kernel Hough transform [93] is used. Here a function (usually gaussian) is distributed over the accumulator array, centred on an indexed reference point. For this work,

the update function is configured for each point that defines a given object model, with the deformation extrema of the object reference point (centroid) defining the function's spatial parameters and gaussian distribution.

Applying the conventional "top hat" approach to an input where no deformation is present produces good results. This can be seen for test image e200 in figure 6.4(a), where only a single maximum is present at the location of the eye object. However, when deformation is present, a significant number of false maxima exist and no one maximum can easily be identified as the object's true location. This is illustrated in the results for test images e300 and e501 in figures 6.4(b) and 6.4(c) respectively. Since little random noise is present in the image data, these false peaks result from correlated noise, which is difficult to overcome [68]. When no deformation is present, all points on the object address and thus increment the same reference point in the accumulator array, resulting in the single maximum shown in figure 6.4(a). Indeed, correlated noise also causes false maxima to be produced, but they are not shown because their values are negligible in the normalised output of figure 6.4(a). Figure 6.5 shows the same result from figure 6.4(a), but a logarithmic, not linear display transfer function is used. When deformation is present, the accumulator array points referenced from each object point rarely correspond. This makes the true location indistinguishable from false locations. The use of the compound (wasp head) object model, as shown in figure 5.23 goes a long way to segmenting the search space, but a significant number of maxima still remain, as shown in figure 6.6(b).

The smooth kernel Hough transform distributes the update in the accumulator array. Palmer *et al.* [93] use a gaussian-like kernel to locate line structures and show this to be optimal, where statistical approaches such as least-squares error are more sensitive to surrounding noise. Statistically, the Hough transform reduces to a maximum likelihood function [126], being equivalent to the log of the probability density function. Hence the accumulator array contains an optimised localisation of the objects of interest [68]. A similar approach is adopted here, only the deformation of the object is used to configure the distribution (gaussian width) and spatial parameters of the accumulator array update function. The results in figure 6.8 show that possible object centroids are represented by localised regions with a large response, not individual maxima. This is because the update functions are more likely to overlap than individual points under conditions of local object deformation. The number of false locations is also significantly reduced, as shown in the result for the bottom of the head sub-structure in figure 6.9. This narrows down the size of the respective landmark neighbourhoods and as such reduces landmark ambiguity. As with the elastic graph matching results, it is important that sufficient deformation is encoded in the training set, otherwise the localisation of object centroids degrades to fixed template matching. This is shown in figures 6.11, 6.12 and 6.13 in chapter 6, where the localisation of the eye sub-structure improves for each test image as the number of training instances increases. Figure 6.14 in chapter 6 shows that fewer instances

reduces the smooth update kernel to a single point and to the conventional "top hat" function described above. The selected object location, an example of which is shown in figure 6.10 does not necessarily correspond with the true object centroid. This results from the deformation present in the input and the deformation encoded in the stored object model. However, this displacement is only slight, and does not affect the resulting neighbourhoods in which the respective landmark points are sought.

9.1.6. Taxonomic Landmark Localisation

As noted above, there are two types of landmark, points on wasp head sub-structure boundaries and whole objects themselves. The latter correspond to the circle landmarks C1 and C2 shown in appendix 2. Due to contour erosion from the filtering of textural clutter, some instances of the circle feature C1 are not well defined in the dominant curve point map, as shown in figure 7.3. As a result, the localisation of this object using the smooth kernel Hough transform and the dominant curve point map will not work, as the results in figures 7.1 and 7.2 show. Since the circles are analytical shapes, a simplified Hough transform is calculated using the more complete edge information at larger scales of analysis in the Mallat wavelet transform. Figure 7.4 shows that good localisation results can be obtained for both circle landmarks. However, the landmarks can be offset in some cases. This is because of false maxima that can emerge in the deformation region of the respective landmark, as shown in figure 7.5. Such false maxima emerge because the circle features and surrounding textural clutter are inseparable in scale space. Since the spatial relationships between the landmarks and centre of gravity for the wasp head are not stable, no precedence can be given to any particular maximum in the deformation region. As such, we cannot easily eliminate false landmark points. This highlights the problem of correlated image noise as discussed by Leavers [68]. Despite this, the results in appendix 8 show that a localisation accuracy of 97% can be obtained for the circle landmark features.

Localising landmark points on the wasp head sub-structures requires that we calculate, for all candidate points in a given neighbourhood, the likelihood of corresponding to the respective landmark. Using the smooth kernel Hough transform to localise the landmark points, where the reference point becomes the landmark and not the object centroid is unsuitable. This is because object deformation results in the displacement of the maximum which corresponds to the reference point, as shown in figure 6.10. As a result, the accuracy of the localised landmark points will significantly affected. The first experiments located landmark points using the local structural match and wavelet edge response criteria. The experiments were performed on the dominant curvature point maps created with the texture filter ROI curve defined in (22). The results show that these criteria are insufficient as false maxima that match the local object structure and have a good wavelet response can exist in the neighbourhood of

each landmark. Figure 7.7 shows how the resulting landmarks are displaced and the spatial relationships between certain landmark pairs is not maintained, resulting in a localisation accuracy of 25%. The second set of experiments augmented the above criteria with the spatial constraints shown in figure 7.10. Here landmark pairs on the eye, top of head and bottom of head have to maintain their spatial orientation. Distance is not relevant because this can fluctuate significantly as a result of natural deformation. For the landmarks on the top and bottom of the head, symmetry with respect to the centre of gravity was also considered. However, although the spatial relationships between the final landmark points are retained, the results in figure 7.11 show that false maxima can also retain these spatial relationships, resulting in a localisation accuracy of 31%. False landmarks exist because of correlated noise in the dominant curvature point map which gives rise to structural ambiguity. This results from residual textural clutter that follows the contour of the head, as shown in figure 7.8. As a result, the local matching and spatial relationship criteria are not suitable for landmark localisation. To improve the results, it has been shown that filtering textural clutter is critical. Figures 7.12 and 7.13 show the landmark localisation results using the above criteria and texture filter ROI curves {16,32,48,64} and {32,64,128,128} respectively. As the ROI increases in size, especially at larger scales of analysis, so the landmarks tend to their true location. The texture filtering results in figure 4.9 show that the texture filter ROI must be sufficiently large at the high end of the scale range because texture features reside at larger scales in the wasp head images. This shows that a direct relationship exists between the calculation of the dominant curvature point map through texture filtering and the accuracy of the landmark points as highlighted in chapter 1. The latter ROI curve produced the best results, with an overall landmark localisation accuracy of 95% being obtained. However, filtering textural clutter also erodes salient contour features, resulting in the displacement of the landmark at the bottom of the eye, as shown for image e100 in appendix 8. Therefore it is important that a balance is maintained, minimising both textural clutter and contour erosion. These findings represent the main contribution to knowledge for this work. Using this technique in conjunction with the smooth kernel Hough transform, the third goal set out in chapter 1 has been achieved.

9.1.7. Species Classification

Using the landmarks shown in appendix 8, the final species classification process is implemented using a back propagation neural network. A set of 5 experiments are carried out, each with a varying number of training epochs, hidden nodes and instances in each training set. The results summarised in appendix 9 show that the greater the number of hidden nodes, the better the classification accuracy of the network. This is due to the significant amount of overlap which exists between some of the species within the taxonomic feature space, in particular species 2, species 3 and species 5. The greater the number of hidden nodes, the

better the delineation of species clusters in the taxonomic feature space. However a balance must be found because if too many hidden nodes are used then the generalisation capability of the network degrades. From these experiments, a classification accuracy of 91% has been obtained. This is a significant improvement over the 65% classification accuracy obtained by experts in the taxonomy of parasitic wasps. Wasp head instances that are misclassified were shown to be both statistically and visually very similar. However, those wasp head instances where the landmarks are displaced are correctly classified. Although the displacement is minimised, it shows that the back propagation network is robust to errors in landmark localisation, allowing wasp head images to be correctly classified. This allows the fourth goal set out in chapter 1 to be achieved.

9.2. Conclusions

The goal of this work was to devise a system to automate the classification of species of parasitic wasp through the automatic localisation of taxonomic landmark points. The results discussed above demonstrate that both of these aims have been achieved, with a landmark localisation accuracy of 95% and a 91% species classification rate being obtained. From these results the following conclusions can be drawn.

(i) Multiscale edge representations derived from gaussian based operators are important for taxonomic landmark localisation and subsequent wasp head classification. Local features extracted from the raw image data are unsuitable because of the textural variation between each wasp head instance and the unstable spatial relationships between each taxonomic landmark point. Texture analysis approaches, including optimised filter bank approaches, which one would intuitively apply to the wasp head images, result in poor edge localisation and the introduction of correlated noise with the identification of false regions. Such noise has been shown to seriously affect the landmark results, reducing the localisation accuracy to 25%. Multiscale edge representations do not introduce erroneous detail at larger scales and the inherent displacement of edges can be overcome by tracking edge events across scale space to their true location at the smallest scale of analysis.

(ii) Given the spatial instability of the taxonomic landmark points and the presence of textural clutter in the dominant curvature point map, we cannot eliminate false landmark locations using the statistical localisation techniques alone. It is therefore critical that textural clutter be filtered out when creating the dominant curvature point map and that a balance is found between filtering textural clutter and minimising salient contour erosion by configuring the texture filter ROI curve appropriately. As such a direct relationship exists between the filtering of textural clutter in creating the dominant curvature point map and the accuracy of the resulting landmark points, with a 95% localisation accuracy being obtained when the appropriate ROI curve is used. This, along with the texture filtering technique described in chapter 4, represent the original contribution to knowledge for this work.

(iii) Contour erosion and false maxima in the Hough parameter space give rise to displaced landmark points. However, the back propagation neural network used for species classification is robust to such errors and can accurately classify a wasp instance when landmark displacement is present. A recognition rate of 91% has been obtained which is a significant improvement over the 65% recognition rate obtained by experts in the taxonomy of parasitic wasps.

Chapter 10

Future Research

10.1. Introduction

The work described throughout this thesis has focused on the development of image feature extraction techniques to facilitate the localisation of taxonomic landmark points on the heads of parasitic wasps for subsequent species classification. The techniques have been designed with a generic use in mind, although the experiments described in this work have focused on the wasp head images in appendix 1. Throughout this work, problems and limitations have been identified with certain techniques, but have not been investigated further because they go beyond the landmark localisation and wasp classification problem domains. This chapter looks at how these techniques can be used in other problem areas and looks at the limitations identified and considers future work to overcome them.

10.2. Object Representation

The shape representation adopted here allows objects to be described at multiple levels of detail and the deformation between object instances to be encoded. A statistical model of deformation is used that is not based on the theory of deformable spline curves (snakes). This is because the textural clutter in the input image would result in poor landmark localisation. The representation used here is based on a dominant curvature point map which contains minimal textural clutter to overcome these problems. The representation itself is generic and can be applied to many different problem domains. Since deformation can be encoded for object localisation with the smooth kernel Hough transform, the representation lends itself to other problems involving biological image data. For applications where the objects have rigid forms, the representation is also useful because only one instance needs to be stored for each object and the redundancy inherent in the representation can be eliminated as a result.

Although the representation is generic, the mechanism used to bind instances of a shape is not. An important constraint imposed on the representation was that it had to conform to the centroidal profile model. This means that only one object point can reside on each radial line from the object's reference point. The mapping of points between two instances therefore does not consider cases where there is more than one point on a radial line. Figure 10.1 illustrates this problem, showing how the polar representation can become complex when the centroidal profile model is not adhered to. This problem arises because we cannot represent shapes as idealised 1D curve functions. This is because shapes are represented as fragmented contour segments in the edge map and we cannot assume idealised input data exists where objects can be extracted in their entirety. To overcome this problem and to extend the representation to other problem domains where shapes with deformable properties do not satisfy the centroidal profile constraint, it would be necessary to find appropriate correspondences between each

curve segment. This problem was not addressed earlier because the wasp head sub-structures satisfy the centroidal profile constraint.

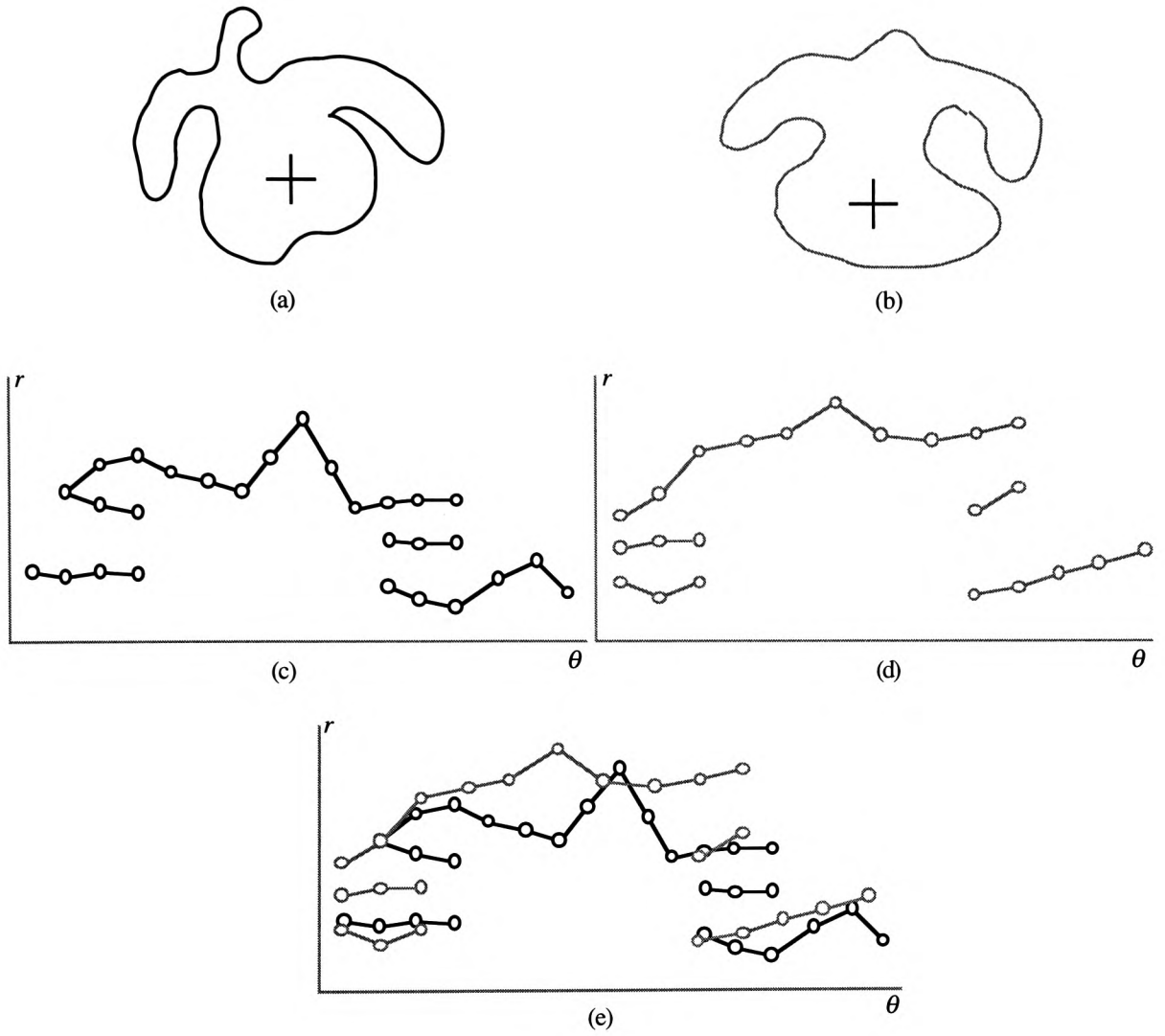


Fig 10.1. (a) and (b) show two example shapes with significant deformation between them. (c) and (d) show the respective polar representations. Since the shapes do not satisfy the centroidal profile constraint, curve segments split and merge along the orientation axis, resulting in many contour segments existing concurrently. (e) shows the two representations overlaid. Binding on both the orientation and radial axes is now more difficult given the fragmented curve segments.

10.3. Global Object Transformations

This work has focused on the localisation of objects under conditions of severe local deformation. In order to apply the object localisation techniques adopted here to a wider range

of applications, it would be necessary to make them robust to global rotation and scaling. Translation invariance does not need to be considered because this is inherent in the object representations used, as described in chapter 5.

Incorporating the global rotation and scaling transformations into Hough based techniques is not difficult, and is widely used throughout the literature [122]. This is done by augmenting the parameter space to include dimensions which represent the rotation and scaling transformations. As well as locating each object in the 2D (image) space described in chapter 6, each possible object rotation and scale would also have to be considered. Maxima which represent possible object locations would then have to be sought in all of the "images" along the rotation and scale axes. The polar representation adopted lends itself to this methodology, as rotation can be implemented by translating the object along the orientation axis and scaling can be implemented by translating the object along the radial axis. Figure 10.2 shows how an object would be localised for rotation invariance. The same principles also apply for scale invariance.

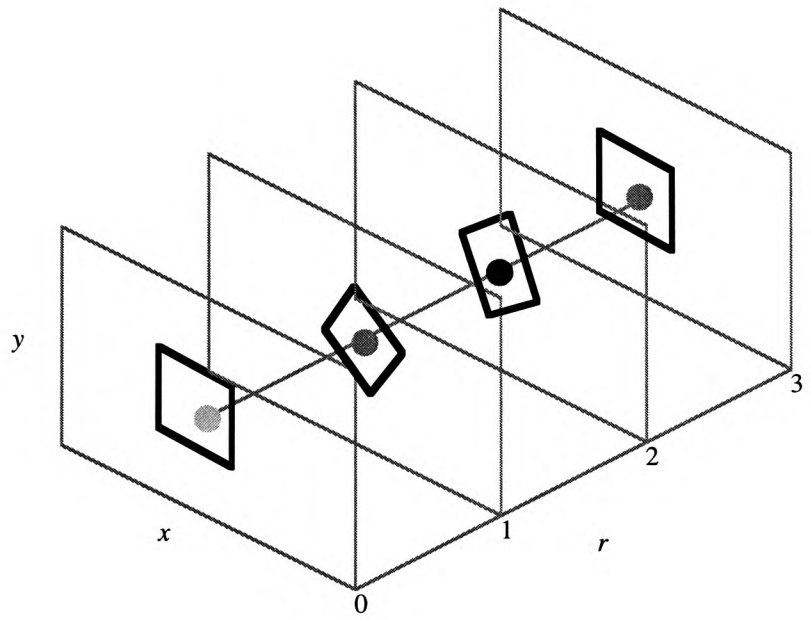


Fig 10.2. To locate a square rotated at 60 degrees in an input image, we need to consider all possible rotations of the square in the parameter space. To locate the object, we need to look for maxima in each image along the rotation axis. The maximum at $r=2$ is the largest and thus represents the location of the square.

However, increasing the size of the parameter space also increases the computational requirements of the Hough transform, in terms of both storage and execution time. When the parameter space was restricted to the image plane, the Hough transform only needed to be calculated at ωh points, where ω and h represent the width and height of the parameter space respectively. By incorporating the rotation and scale axes, the Hough transform would have to

be calculated at ωhrs points, where r and s represent the number of discrete points along the rotation and scale axes respectively.

10.4. Contour / Texture Delineation

This work describes a novel technique to recombine the Mallat wavelet transform and filter textural clutter leaving only salient contour features. In deriving an intermediate representation to localise taxonomic landmarks on the wasp head, it was shown that configuring the texture filter ROI curve at each scale was critical, eliminating the structural ambiguity inherent in the deformable input data. The experiments described in this work also consider the application of this technique to other images, including simple synthetic images and SEM images of the heads of Fruitflies *(Drosophila melanogaster)*. This technique can thus be applied to other problem domains, provided the ROI curve is configured appropriately. However, one important problem that must be overcome concerns the calculation of local wavelet activity around a given edge point. It was shown that two edge events which have a large response and lie in close proximity within the wavelet transform produce a large interaction result. This makes it difficult to differentiate between salient structures of small spatial extent and regions of significant textural activity. This is an important problem that must be addressed in future research.

10.5. Future Research

The above sections have considered how the techniques described throughout this work can be applied to other problem domains and discussed future work that is necessary to extend the applicability of the techniques further. This section discusses two important areas of future research. The first involves texture segmentation. It was shown that due to the variance in textural properties, false regions are created in the final edge map. The presence of such clutter was shown to produce poor landmark localisation results. Coupled with the problem of poor edge localisation, texture analysis was deemed unsuitable for locating the taxonomic landmarks in the wasp head images. One of the major problems is that texture segmentation is based on a multi-dimensional feature space representing the local image intensity, spatial frequency and orientation. Considering these features together gives rise to the problem of false regions. It is hypothesised that regions, or whole shapes can exist within a sub-space of this texture feature space. If we consider the segmentation result in figure 3.1, we can see that the eye object has good segmentation results in the spatial frequency feature image but poor segmentation results in the orientation feature image. Using the feature images in figures 3.1(a) and 3.1(b) to create the final texture region segmentation results in the poor edge map shown in figure 3.1(d). However, we could consider the eye object to exist only along the spatial frequency axis of the

texture feature space, ignoring the orientation information. This would result in better segmentation, but further research is required to determine how high level knowledge about the eye and other objects can be used to guide search in the texture feature space and how this can be applied to other problem domains.

The second important area for future research involves the classification of species using a neural network architecture. Although the focus of this work has been to locate the taxonomic landmark points under conditions of severe local deformation, it is important to extend the classification system described here so it can be used by taxonomists in the field. Given the vast number of species available, the network architecture would need to be extended to incorporate these. One of the issues discussed in chapter 1 was that only about 25% of species have been identified. To address this problem, rejection can be built into the network architecture. This involves determining when the input does not match any of the known species. This would be an important area of research because taxonomists could use the system to identify possible new species.

References

1. Adams, R., Bischof, L. "Seeded Region Growing". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 16, No. 6, pp. 641-647. 1994.

2. Ahuja, N. "A Transform for Multiscale Image Segmentation by Integrated Edge and Region Detection". *IEEE Trans on Pattern Analysis and Machine Intelligence*. Vol. 18, No. 12, pp. 1211-1235. 1996.

3. Akra, M. *et al.* "Sampling of Images for Efficient Model Based Vision". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 21, No. 1, pp. 4-11. 1999.

4. Anderson, I.M., Bezdek, J.C. "Curvature and Tangential Deflection of Discrete Arcs: A Theory Based on the Commutator of Scatter Matrix Pairs and its Application to Vertex Detection in Planar Shape". *IEEE Trans on Pattern Analysis and Machine Intelligence*. Vol. 6, No. 1, pp. 27-40. 1984.

5. Angel, P.N., Morris, C.W. "A Neural Network Pipeline Architecture for Image Feature Extraction". *1st Int. Conf. on Imaging Science, Systems and Technology*. ed. Arabnia, H.R. pp. 311-318. 1997.

6. Attneave, F. "Symmetry, Information and Memory for Patterns". *American Journal of Psychology*. Vol. 68, pp. 209-222. 1955.

7. Ballard, D.H. "Generalising the Hough Transform to Detect Arbitrary Shapes". *Pattern Recognition*. Vol. 13, pp. 111-122. 1981.

8. Becker, S., Plumbley, M. "Unsupervised Neural Network Learning Procedures for Feature Extraction and Classification". *Applied Intelligence*. Vol. 6, pp. 185-203. 1996.

9. Bhandarkar, S.M. *et al.* "A Hierarchical Neural Network and its Application to Image Segmentation". *Mathematics and Computers in Simulation*. Vol. 41. pp. 337-355. 1996.

10. Blaszka, T., Deriche, R. "Recovering and Characterising Image Features Using an Efficient Model Based Approach". Technical Report. No. 2422. INRIA - Sophia Antipolis. 1994.

11. Bookstein, F.L. "Shape and the Information in Medical Images : A Decade of Morphometric Synthesis". *Computer Vision and Image Understanding*. Vol. 66, No. 2. pp. 97-118. 1997.

12. Brown, M.B. "Inherent Bias and Noise in the Hough Transform". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. PAMI-5, 1983.

13. Canny, J. "A Computational Approach to Edge Detection". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. PAMI-8, pp. 679-714. 1986.

14. Carpenter, G.A., *et al.* "The What-and Where Filter". *Computer Vision and Image Understanding*. Vol. 69, No. 1, pp.1-22. 1999.

15. Carragher, B., Smith, P.R. "Advances in Computational Processing for Microscopy". *Journal of Structural Biology*. Vol. 116, pp. 2-8. 1996.

16. Chen, Y.Q.*et al.* "Texture Classification Using Statistical Geometric Features". *Proc. IEEE*. pp. 446-450. 1994.

17. Chen, H., Liu, R. "An On-line Unsupervised Learning Machine for Adaptive Feature Extraction". *IEEE Trans. on Circuits and Systems II : Analogue and Digital Signal Processing*. Vol. 41, No. 2, pp.87-97. 1994.

18. Cho, K., Dunn, S.M. "Learning Shape Classes". *IEEE Trans Pattern Analysis and Machine Intelligence*. Vol. 16, No. 9, pp. 882-887. 1994.

19. Cohen, A., Kovacevic, J. "Wavelets : The Mathematical Background". *Proc. IEEE*. Vol. 84, No. 4, pp. 514-522. 1996.

20. Dawah, H.A. *et al.* "Structure of the Parasitoid Communities of Grass Feeding Chalcid Wasps". *Journal of Animal Ecology*. Vol. 64, pp. 708-720. 1995.

21. Daly, H.V. "Insect Morphometrics". *Ann. Rev. Entomology*. Vol. 30, pp. 415-438. 1985.

22. Daubechies, I. "Where Do Wavelets Come From? - A Personal Point of View". *Proc. IEEE*. Vol. 84, No. 4, pp. 510-513. 1996.

23. Daugman, J.G. "Complete Discrete 2D Gabor Transforms by Neural Networks for Image Analysis and Compression". *IEEE Trans. on Acoustics, Speech and Signal Processing*. Vol. 36, No. 7, pp. 1169-1179. 1988.

24. Davis, L.S. "Hierarchical Generalised Hough Transforms and Line Segment Based Generalised Hough Transforms". *Pattern Recognition.* Vol. 15, No. 4, pp. 277-285. 1982.

25. Davis, L.S., Mitiche, A. "Edge Detection in Textures". *Computer Graphics and Image Processing.* Vol. 12, pp. 25-39. 1980.

26. Duc, B. *et al.* "Face Authentication with Gabor Information on Deformable Graphs". *Proc. IEEE. Trans. on Image Processing.* Vol. 8, No. 4, pp.504-516. 1999.

27. Dudek, G., Tsotsos, J.K. "Shape Representation and Recognition from Curvature". *Proc IEEE.* pp. 35-41. 1991.

28. Dudek, G., Tsotsos, J.K. "Shape Representation and Recognition from Multiscale Curvature". *Computer Vision and Image Understanding.* Vol. 68, No. 2, pp.170-189. 1997.

29. Dunn, S.M. *et al.* "Interpreting Transmission Electron Micrographs of Insect Hemocytes". *Journal of Computer Assisted Microscopy.* Vol. 2, No. 3, pp. 133-160. 1990.

30. Earnshaw, G. "A Review of Image Preprocessing Techniques for Artificial Neural Networks". Technical report. Centre for Intelligent Systems, University of Plymouth. 1993.

31. Earnshaw, G. "Investigation of Texture Based Segmentation using Gabor Filters". Technical report. Centre for Intelligent Systems, University of Plymouth. 1993.

32. Fausett, L. "Fundamentals of Neural Networks - Architectures, Algorithms and Applications". Prentice Hall International, Inc. 1994.

33. Fidrich, M., Thirion, J-P. "Multiscale Extraction and Representation of Features from Medical Images". Technical Report. No. 2365. INRIA - Sophia Antipolis. 1994.

34. Fidrich, M., Thirion, J-P. "Stability of Corner Points in Scale Space: The Effects of Small Nonrigid Deformations". *Computer Vision and Image Understanding.* Vol. 72, No. 1, pp.72-83. 1998.

35. Foley, D.H. "Considerations of Sample and Feature Size". *IEEE Trans. on Information Theory.* Vol. 18, No. 5, pp. 618-626. 1972.

36. Freeman, H., Davis, L.S. "A Corner Finding Algorithm for Chain Coded Curves". *IEEE Trans. Comput.* Vol. C-26, pp.297-303. 1977.

37. Full, W.E., Ehrlich, R. "Fundamental Problems Associated with Eigenshape Analysis and Similar Factor Analysis Procedures". *Math. Geol.* Vol. 18, No. 15, pp. 451-463. 1986.

38. Ganesan, I., Bhattacharyya, P. "Edge Detection in Untextured and Textured Images - A Common Computational Framework". *IEEE Trans. Systems, Man and Cybernetics.* Vol. 27, pp. 823-834. 1997.

39. Ghosal, S., Mehrotra, R. "Range Surface Characterisation and Segmentation Using Neural Networks". *Pattern Recognition.* Vol. 28, No. 5, pp. 711-727. 1995.

40. Goldstein, M. "K-Nearest Neighbor Classification". *IEEE Trans. on Information Theory.* Vol. 18, No. 5, pp. 627-630. 1972.

41. Gokmen, M., Jain, A.K. "$\lambda\tau$-space Representation of Images and Generalised Edge Detector". *IEEE Trans. Pattern Analysis and Machine Intelligence*. Vol. 19, No. 6, pp. 545-563. 1997.

42. Gordon, I.E. "Theories of Visual Perception - 2nd Edition". Wiley, Chichester. 1997.

43. Grimson, W.E.L, Huttenlocher, D.P. "On the Sensitivity of the Hough Transform for Object Recognition". *IEEE Trans. Pattern Analysis and Machine Intelligence.* Vol 12, No 3, pp. 255-274. 1990.

44. Hayakawa, H. *et al.* "A Computational Model for Shape Estimation by Integration of Shading and Edge Information". *Neural Networks.* Vol. 7, No. 8, pp. 1193-1209. 1994.

45. Hoffman, D.D., Richards, W.A. "Parts of Recognition". *Cognition.* Vol. 18. pp.65-96. 1984.

46. Horn, B.K.P. "Robot Vision". MIT Press. 1986.

47. Hough, P.V.C. "Method and Means for Recognising Complex Patterns". US Patent No. 3069654, 1962.

48. Hsieh, J-W, *et al.* "Image Registration Using a New Edge-Based Approach". *Computer Vision and Image Understanding.* Vol. 67, No. 2, pp. 112-130. 1997.

49. Hubel, D., Wiesel, T. "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex". *J. Physiol. (London).* Vol. 160, pp. 106-154. 1962.

50. Iftekharuddin, K.M., Jemili, K. "Feature-Based Neural Wavelet Optical Character Recognition System". *Optical Engineering*. Vol. 34, No. 11, pp. 3193-3199. 1995.

51. Illingworth, J., Kittler, J. "The Adaptive Hough Transform". *IEEE Trans. Pattern Analysis and Machine Intelligence*. Vol. 9, No. 5, pp. 690-698. 1987

52. Intrator, N. *et al.* "Face Recognition Using a Hybrid Supervised / Unsupervised Neural Network". *Pattern Recognition Letters*. Vol. 17, pp. 67-76. 1996.

53. Jang, J., Shin, D. "Parallel Optical-Feature Extraction by use of Rotationally Multiplexed Holograms". *Optics Letters*. Vol. 21, No. 19, pp. 1612 - 1614. 1996.

54. Jolion, J.M., Montanvert, A. "The Adaptive Pyramid : A Framework for 2D Image Analysis". *CVGIP:Image Understanding*. Vol. 55, No. 3, pp. 339-348. 1992.

55. Kapouleas, I., Kulikowski, C. " A Model Based System for the Interpretation of MR Human Brain Scans". *Medical Imaging II*, SPIE Vol. 914, pp. 429-437. 1988.

56. Kass, M. *et al.* "Snakes : Active Contour Models". *Proc. IEEE*. pp. 259-268. 1987.

57. Kass, M. *et al.* "Snakes : Active Contour Models". *Int. Journal of Computer Vision*. pp. 321-331. 1988.

58. Kirsch, R. "Computer Determination of the Constituent Structure of Biological Images". *Comput. Biomed.* Vol. 4. pp. 315-328. 1971.

59. Koenderink, J.J. "The Structure of Images", *Biolog. Cybern.* Vol. 50, pp. 363-370. 1984.

60. Kohonen, T. "Self-Organisation and Associative Memory", 2nd Edition. Springer, Berlin, 1988.

61. Kreho, A. *et al.* "Computer Assisted Feature Extraction for Dolphin Identification". *First Int. Conf. on Imaging Science, Systems and Technology (CISST)*. pp. 440-445. 1997.

62. Kulkarni, A.D. "Artificial Neural Networks for Image Understanding". Van Nostrand Reinhold. 1993.

63. Lades, M. *et al.* "Distortion Invariant Object Recognition in the Dynamic Link Architecture". *IEEE Trans on Computers.* Vol. 42, No. 3, pp. 300-310. 1993.

64. Lai, K., Chin, R.T. "Deformable Contours : Modelling and Extraction". *IEEE Trans. Pattern Analysis and Machine Intelligence.* Vol. 17, No. 11, pp. 1084-1090. 1997.

65. Lai, K., Chin, R.T. "On Classifying Deformable Contours Using the Generalised Active Contour Model". *Third Int. Conf. on Automation, Robotics and Computer Vision.* pp.930-934. 1994.

66. Lam, K., Yan, H. "An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View". *IEEE Trans. on Pattern Analysis and Machine Intelligence.* Vol. 20, No. 7, pp. 673-686. 1998.

67. Leavers, V.F. "Shape Detection in Computer Vision Using the Hough Transform". Springer Verlag , London, 1992.

68. Leavers, V.F. "Which Hough Transform". *CVGIP:Image Understanding.* Vol. 58, No. 2, pp.250-264. 1993.

69. Lin, W-G., Wang, S-S. "A New Neural Model for Invariant Pattern Recognition". *Neural Networks.* Vol. 9, No. 5, pp. 899-913. 1996.

70. Lindeberg, T. "Scale-Space for Discrete Signals". *IEEE Trans. on Pattern Analysis and Machine Intelligence.* Vol. 12, No. 3, pp. 234-254. 1990.

71. Liu, K. *et al.* "Algebraic Feature Extraction for Image Recognition Based on an Optimal Discriminant Criterion". *Pattern Recognition.* Vol. 26, No. 6, pp. 903-911. 1993.

72. Lohmann, G.P. "Eigenshape Analysis of Microfossils : A General Morphometric Procedure for Describing Changes in Shape". *Math Geol.* Vol. 15, No. 6, pp. 659-672. 1983.

73. Low, A. "Introductory Computer Vision and Image Processing". McGraw Hill. 1991.

74. Lowe, D.G. "Perceptual Organisation and Visual Learning". Kluwer Academic. 1985.

75. Lu, Y., Jain, R.C. "Reasoning about Edges in Scale Space". *IEEE Trans. on Pattern Analysis and Machine Intelligence.* Vol. 14, No. 4, pp. 450-468. 1992.

76. Lu, S.W., Xu, H. "Textured Image Segmentation Using Autoregressive Model and Artificial Neural Network". *Pattern Recognition.* Vol. 28, No. 12, pp.1807-1817. 1995.

77. Mallat, S. "Wavelets for a Vision". *Proc. IEEE.* Vol. 84, No.4, pp. 604-614. 1996.

78. Mallat, S., Hwang, W.L. "Singularity Detection and Processing with Wavelets". *IEEE Trans. on Information Theory.* Vol. 38, No. 2, pp. 617-643. 1992.

79. Mallat, S., Zhong, S. "Characterisation of Signals from Multiscale Edges". *IEEE Trans. Pattern Analysis and Machine Intelligence.* Vol. 14, No. 7, pp.710-732. 1992.

80. Mandelbrot, B.B. "The Fractal Geometry of Nature". W.H. Freeman, San Francisco, 1982.

81. Marr, D. "Vision". San Francisco: W.H. Freeman. 1982.

82. Marr, D., Hildreth, E. "Theory of Edge Detection". *Proc. Royal Soc. London*, Vol. 207, pp.187-217, 1980.

83. Marsic, I. "Evidential Reasoning In Visual Recognition". *Intelligent Engineering Systems through Artificial Neural Networks.* Vol. 4.

84. Marston, R.E., Shih, J.C. "Polygonal Approximation of Outlines by Scale Based Dominant Point Detection". *Proc. IEE Image Processing And Its Applications*, pp. 350-354. 1995.

85. Meer, P. *et al.* "Extraction of Trend Lines and Extrema from Multiscale Curves". *Pattern Recognition.* Vol. 21, No. 3, pp. 217-226. 1988.

86. Mehrotra, R., *et al.* "Gabor Filter-Based Edge Detection". *Pattern Recognition.* Vol. 25, No. 12, pp. 1479-1494. 1992.

87. Moghaddam, B., Pentland, A. "Face Recognition Using View-Based and Modular Eigenspaces". *Proc. SPIE.* Vol. 2277. pp.12-21. 1994.

88. Moghaddam, B., Pentland, A. "Probabilistic Visual Learning for Object Recognition". *Proc. IEEE.* pp. 786-793. 1995.

89. Mokhtarian, F., Mackworth, A.K. "A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves". *IEEE Trans. on Pattern Analysis and Machine Intelligence.* Vol. 14, No. 8, pp.789-805. 1992.

90. Mui, J. *et al.* "Feature Selection in Automated Classification of Blood Cell Neutrophils". *Proc. IEEE.* pp. 486 - 491. 1978.

91. Niessen, W.J. *et al.* "Nonlinear Multiscale Representations for Image Segmentation". *Computer Vision and Image Understanding.* Vol. 66, No. 2, pp. 233-245. 1997.

92. Palmer, P.L. *et al.* "A Hough Transform Algorithm with a 2D Hypothesis Testing Kernel". *CVGIP: Image Understanding.* Vol. 58, No. 2, pp. 221-234. 1993.

93. Palmer, P.L. *et al.* "An Optimising Line Finder Using a Hough Transform Algorithm". *Computer Vision and Image Understanding.* Vol. 67, No. 1, pp. 1-23. 1997.

94. Pentland, A. *et al.* "View-Based and Modular Eigenspaces for Face Recognition". *Proc IEEE.* pp. 84-91. 1994.

95. Perry, A., Lowe, D. "Segmentation of Non-Random Textures Using Zero Crossings". *Proc. IEEE.* 1989.

96. Philip, K.P. "Automatic Detection of Myocardial Contours in Cine Computed Tomographic Images". PhD Thesis, University of Iowa, 1991

97. Pichler, O. *et al.* "A Comparison of Texture Feature Extraction Using Adaptive Gabor Filtering, Pyramidal and Tree Structured Wavelet Transforms". *Pattern Recognition.* Vol. 29, No. 5, pp. 733-742. 1996.

98. Porat, M., Zeevi, Y.Y. "The Generalised Gabor Scheme of Image Representation in Biological and Machine Vision". *IEEE Trans. on Pattern Analysis and Machine Intelligence.* Vol. 10, No. 4, pp. 452-467. 1988.

99. Porat, M., Zeevi, Y.Y. "Localised Texture Processing in Vision : Analysis and Synthesis in Gaborian Space". *IEEE Trans. Biomed. Engineering.* Vol. 36, No. 1, pp. 115-129. 1989.

100. Prewitt, J.M.S., Mendelsohn, M.L. "The Analysis of Cell Images". *Annals New York Academy of Sciences.* pp. 1035-1053.

101. Ramalho, M., Curtis, K.M. "Edge Detection Using Neural Network Arbitration". *Proc IEE Image Processing and Its Applications*. pp. 514-518. 1995

102. Randen, T., Husøy, J. "Texture Segmentation Using Filters with Optimised Energy Seperation". *IEEE Trans. on Image Processing*. Vol. 8, No. 4, pp. 571-582. 1999.

103. Randen, T. *et al.* "Optimal Filtering for Unsupervised Texture Feature Extraction". *Proc. SPIE Conf. Visual Communications and Image Processing*. pp. 441-452. 1996.

104. Rattarangsi, A., Chin, R.T. "Scale-Based Detection of Corners of Planar Curves". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 14, No. 4, pp. 430-449. 1992.

105. Ravichandran, G., Casasent, D. "Advanced In-Plane Rotation-Invariant Correlation Filters". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 16, No. 4, pp. 415-420. 1994.

106. Ravichandran, G., Trivedi, M. "Circular-Mellin Features for Texture Segmentation". *IEEE Trans. on Image Processing*. Vol. 4, No. 12, pp. 1629-1640. 1995.

107. Reed, T., Wechsler, H. "Segmentation of Textured Images and Gestalt Organisation Using Spatial/Spatial-Frequency Representations". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 12, No. 1, pp. 325-331. 1990.

108. Reed, T., Wechsler, H. "Spatial/Spatial-Frequency Representations for Image Segmentation and Grouping". *Image and Vision Computing*. Vol. 9, No. 3, pp. 175-193. 1991.

109. Roberts, L.G. "Machine Perception of Three Dimensional Solids". *Optical and Electro-optical Information Processing*. ed. Tippet, J.T. MIT Press, Cambridge, MA. 1965.

110. Rosenfeld, A., Johnston, E. "Angle Detection on Digital Curves". *IEEE Trans. Comput.* Vol. C-22, pp.875-878. 1973.

111. Rosenfeld, A., Weazka, J.S. "An Improved Method of Angle Detection on Digital Curves". *IEEE Trans. Comput.* Vol. C-24, pp.940-941. 1975.

112. Russ, J.C. "The Image Processing Handbook". CRC Press. 1992.

113. Saint-Marc, P., Medioni, G. "Adaptive Smoothing for Feature Extraction". *Proc. Image Understanding*. Vol. 2, pp. 1100-1113. 1988.

114. Sahiner, B. *et al.* "Classification of Mass and Normal Breast Tissue : A Convolution Neural Network Classifier with Spatial Domain and Texture Images". *IEEE Trans. on Medical Imaging*. Vol. 15, No. 5, pp. 598-610. 1996.

115. Sankar, P.V., Sharma, C.V. "A Parallel Procedure for the Detection of Dominant Points on Digital Curves". *Comput. Vision Graphics Image Processing*. Vol. 7, pp.403-412. 1978.

116. Sarkar, S., Boyer, K.L. "Optimal Infinite Impulse Response Zero Crossing Based Edge Detectors". *CVGIP:Image Understanding*. Vol. 54, No. 2, pp. 224-243. 1991.

117. Shang, C., Brown, K. "Principal Features-Based Texture Classification with Neural Networks". *Pattern Recognition*. Vol. 27, No. 5, pp. 675-687. 1994.

118. Shustorovich, A. "Scale Specific and Robust Edge/Line Encoding with Linear Combinations of Gabor Wavelets". *Pattern Recognition.* Vol. 27, No. 5, pp. 713-725. 1994.

119. Shustorovich, A. "A Subspace Projection Approach to Feature Extraction : The Two-Dimensional Gabor Transform for Character Recognition". *Neural Networks*. Vol. 7, No. 8, pp.1295-1301. 1994.

120. Smith, P.R., Gottesman, S.M. "The Micrograph Data Processing Program". *Journal of Structural Biology*. Vol. 116, pp. 35-40. 1996.

121. Sobel, I. "Camera Models and Machine Perception". AIM-21, Stanford AI Lab, Palo Alto.

122. Sonka, M. *et al.* "Image Processing, Analysis and Machine Vision". Chapman & Hall, London, 1993.

123. Srinivasa, N., Jouaneh, M. "A Neural Network for Invariant Pattern Recognition". *IEEE Trans. on Signal Processing*. Vol. 40, pp. 1595-1599. 1992.

124. Srinivasa, N., Jouaneh, M. "An Invariant Pattern Recognition Machine Using a Modified ART Architecture". *IEEE Trans. on System, Man and Cybernetics*. Vol. 5. pp. 1432-1437. 1993.

125. Staib, L.H., Duncan, J.S. "Boundary Finding with Parametrically Deformable Models". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 14, No. 11, pp.1061-1075. 1992.

126. Stephens, R.S. "Probabilistic Approach to the Hough Transform". *Image Vision Computing*. Vol. 9, No. 1, 1991.

127. Straney, D.O. "Median Axis Methods in Morphometrics". --------

128. Strela, V. *et al.* "The Application of Multiwavelet Filterbanks to Image Processing". *IEEE Trans Image Processing*. Vol. 8, No. 4, pp.548-563. 1999.

129. Suetens, P., Oosterlinck, A. "Using Expert Systems for Image Understanding". *Int. Journal of Pattern Recognition and Artificial Intelligence*. Vol. 1, No. 2, pp. 237-250. 1987.

130. Suetens, P. *et al.* "Recognition of the Coronary Blood Vessels on Angiograms Using Hierarchical Model-Based Iconic Search". *Proc IEEE*. pp. 576-581. 1989.

131. Swets., D.L., Weng, J. "Using Discriminant Eigenfeatures for Image Retrieval". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 18, No. 8, pp.831-836. 1996.

132. Talukder, A., Casasent, D. "General Methodology for Simultaneous Representation and Discrimination of Multiple Object Classes". *Optical Engineering*. Vol. 37, No. 3, pp. 904-913. 1998.

133. Tang, X., Stewart, K. "Plankton Image Classification Using Novel Parallel-Training Learning Vector Quantisation Network". *Proc IEEE*. pp. 1227-1236. 1996

134. Teh, C., Chin, R.T. "On the Detection of Dominant Points on Digital Curves". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 11, No. 8, pp.859-872. 1989.

135. Terzopoulos, D., Metaxas, D. "Dynamic 3D Models with Local and Global Deformations : Deformable Superquadratics". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 13, No. 7, pp.703-714. 1991.

136. Thirion, J-P., Gourdon, A. "The 3D Marching-Lines Algorithm : New Results and Proofs". INRIA. 1993.

137. Tsotsos, J.K. "Knowledge and the Visual Process : Content, Form and Use". *Pattern Recognition*. Vol. 17, No. 1, pp. 13-27. 1984.

138. Turk, M., Pentland, A. "Eigenfaces for Recognition". *Journal of Cognitive Neuroscience*. Vol. 3, No. 1, pp. 71-86. 1991.

139. Ulupinar, F., Medioni, G. "Finding Edges Detected by a LoG Operator". *Computer Vision, Graphics and Image Processing*. Vol. 51, pp. 275-298. 1990.

140. Unser, M. "Local Linear Transforms for Texture Measurements". *Signal Processing*. Vol. 11, pp. 61-79. 1986.

141. Unser, M. *et al.* "A Family of Polynomial Spline Wavelet Transforms". *Signal Processing*. Vol. 30. pp. 141-162. 1993.

142. Unser, M. "Texture Classification and Segmentation Using Wavelet Frames". *IEEE Trans. on Image Processing,* Vol. 14, No. 11, pp. 1549-1560. 1995.

143. Vaidyanathan, M., *et al.* "Tumor Volume Measurements Using Supervised and Semi-Supervised MRI segmentation methods". ----------

144. Vernon, D. "Two-dimensional Object Recognition Using Partial Contours". *Image and Vision Computing*. Vol. 5, No. 1, pp. 21-27. 1987.

145. Weeks, P.J.D. *et al.* "Automating the Identification of Insects : A New Solution to an Old Problem". *Bull. Entomological Research*. Vol. 87, pp. 203-211. 1997.

146. Widrow, B., Winter, R. "Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition". *Computer*. Vol. 21, pp. 25-39. 1988.

147. Widrow, B. *et al.* "Layered Neural Nets for Pattern Recognition". *IEEE Trans. on Acoustics Speech and Signal Processing*. Vol. 26, pp. 1109-1118. 1988.

148. Wilkins, M. *et al.* "A Comparison of Some Neural and Non-neural Methods for Identification of Phytoplankton from Flow Cytometry Data". CABIOS, Vol. 12, No. 1, pp. 9-18. 1996.

149. Wilson, R.C. *et al.* "Structural Matching with Active Triangulations". *Computer Vision and Image Understanding*. Vol. 72, No. 1, pp. 21-38. 1998.

150. Witkin, A.P. "Scale Space Filtering". *Proc. Int. Conf. AI.*, Vol. 511. 1983.

151. Xie, X. *et al.* "Corner Detection by a Cost Minimisation Approach". *Pattern Recognition.* Vol. 26, pp. 1235-1243. 1993.

152. Yu, S., Berthod, M. "A Game Strategy Approach for Image Labelling". *Computer Vision and Image Understanding.* Vol. 61, No. 1, pp. 32-37. 1995.

153. Yuille, A.L., Poggio, T.A. "Scaling Theorems for Zero Crossings". *IEEE Trans. on Pattern Analysis and Machine Intelligence.* Vol. PAMI-8, pp. 15-25. 1986.

154. Zhang, X., Deng, H. "Distributed Image Edge Detection Methods and Performance". *Proc. IEEE.* pp. 136-143.

155. Zheng, Y-J. "Feature Extraction and Image Segmentation Using Self Organising Networks". *Machine Vision and Applications.* Vol. 8, pp. 262-274. 1995.

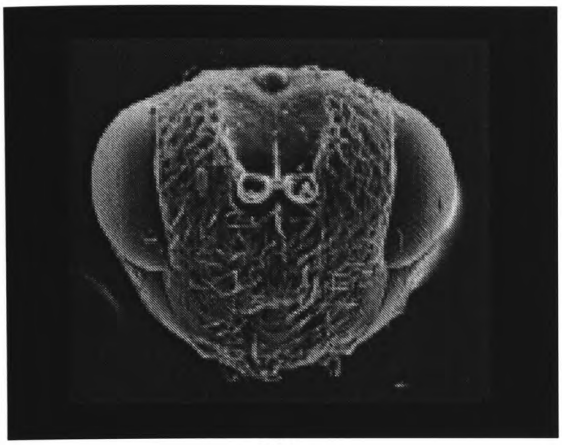
156. Zhu, Y.M., Goutte, R. "Analysis and Comparison of Space/Spatial Frequency and Multiscale Methods for Texture Segmentation". *Optical Engineering*, Vol. 34, No. 1, pp.269-282. 1995.

157. Zhu, Y.M. "On the Use of Two-Dimensional Wigner-Ville Distribution for Texture Segmentation". *Signal Processing.* Vol. 30, pp.329-353. 1993.

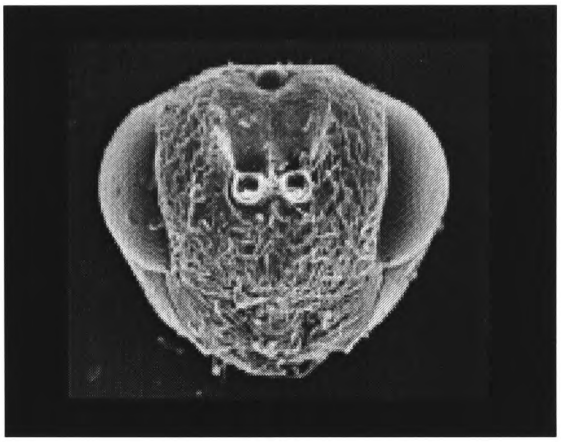
Appendix 1

SEM Image Library

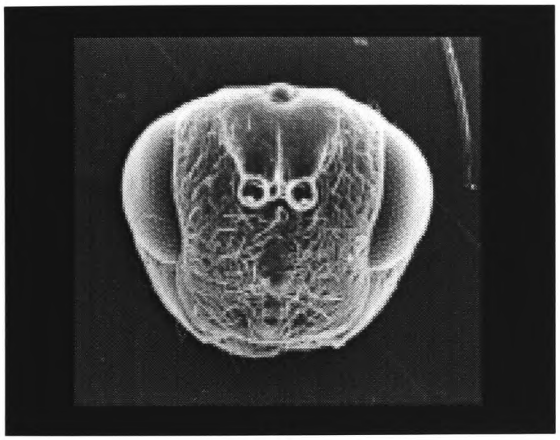
Species 1 : *Tetramesa linearis*



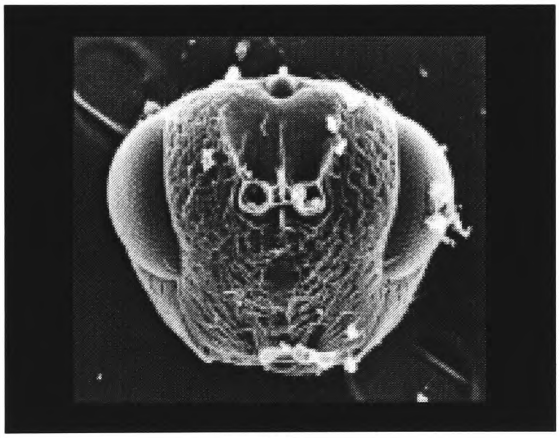
e100



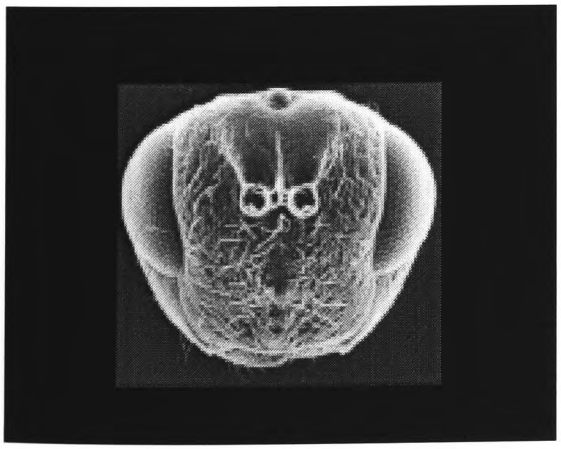
e101



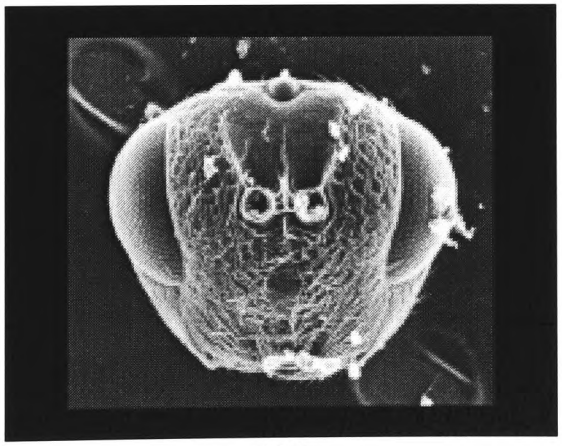
e102



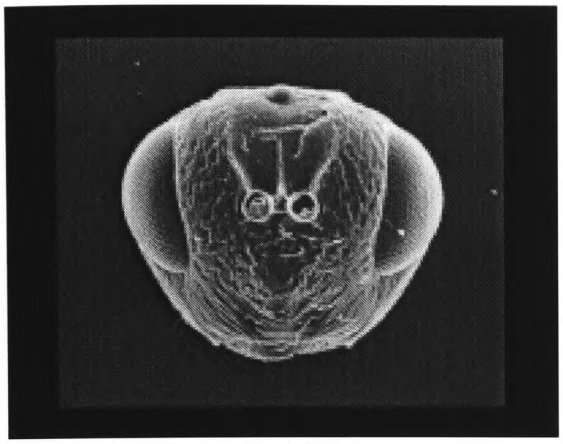
e103



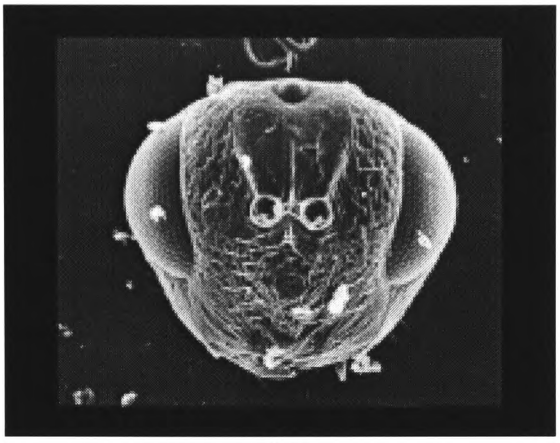
e104



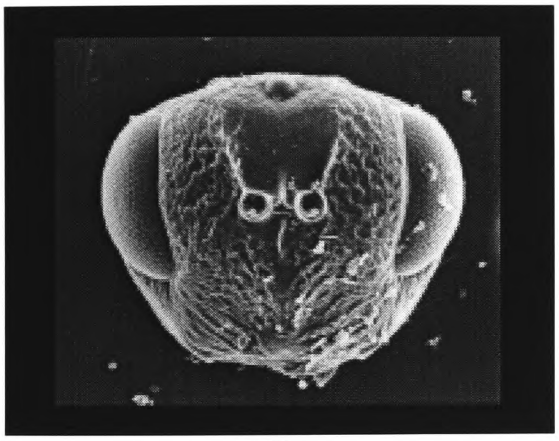
e105



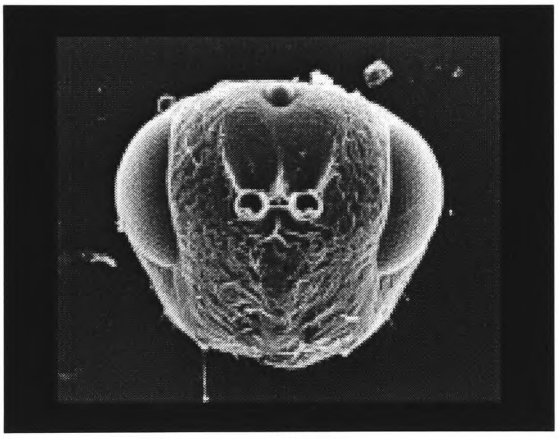
e106



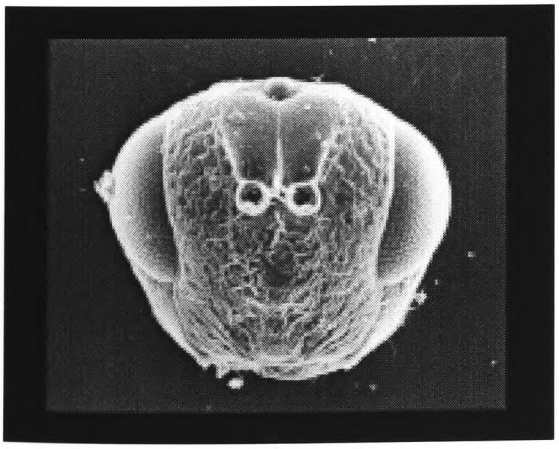
e107



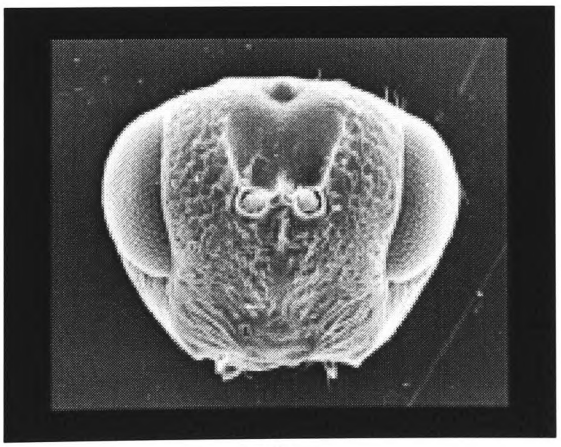
e108



e109

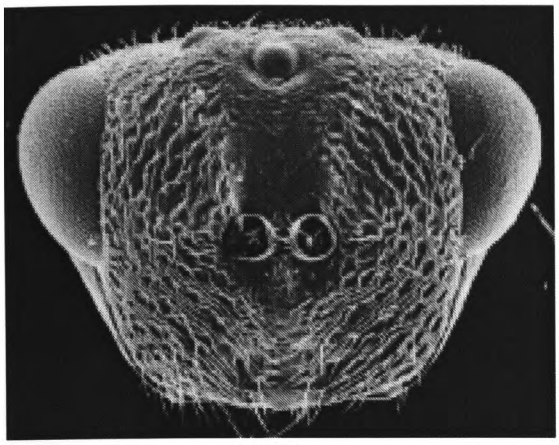


e110

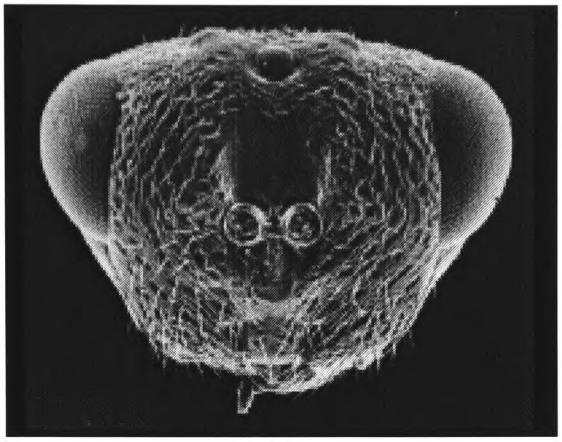


e111

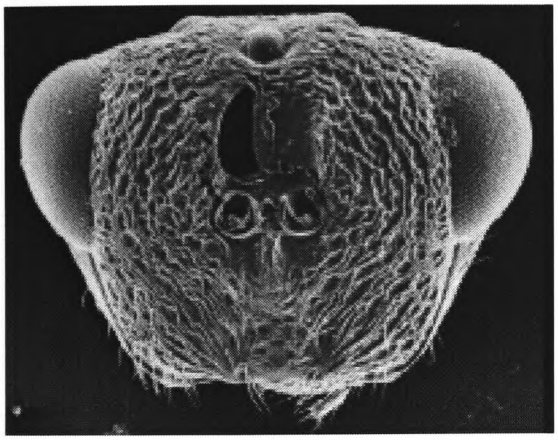
Species 2 : *Tetramesa hyalipennis*



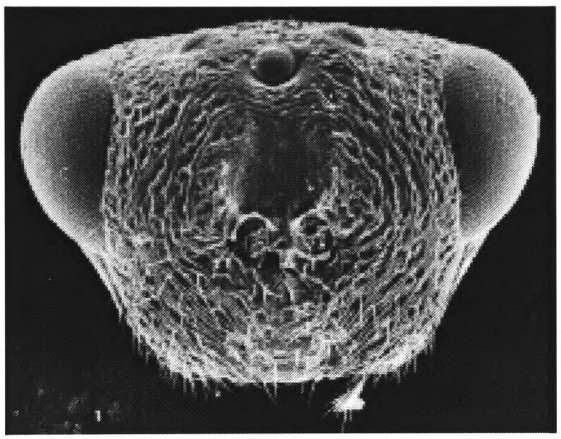
e200



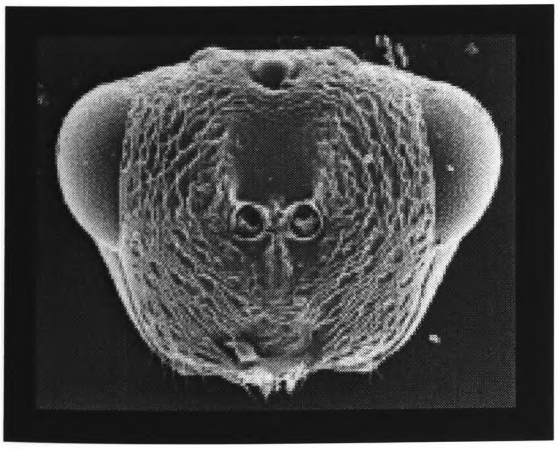
e201



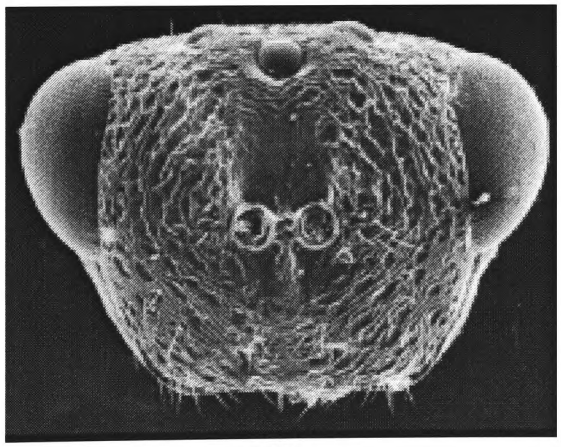
e202



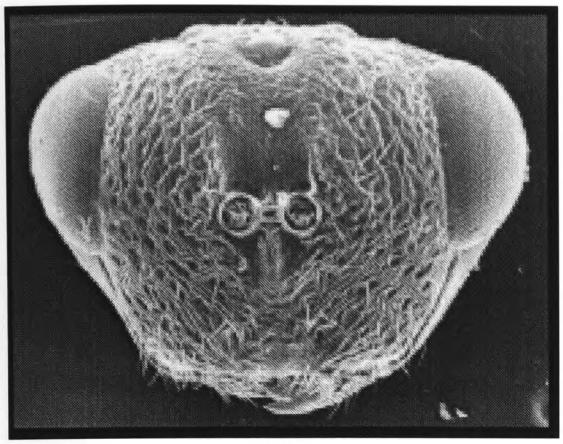
e203



e204

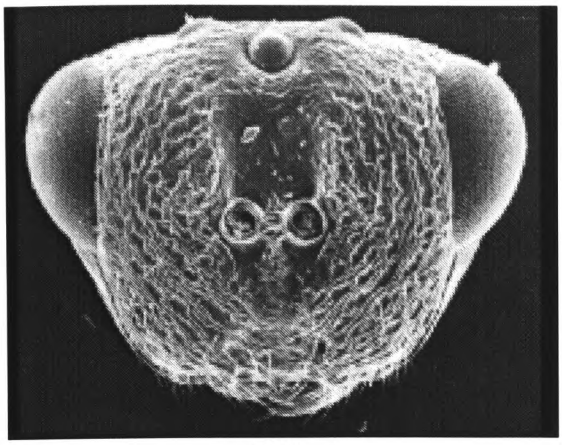


e205

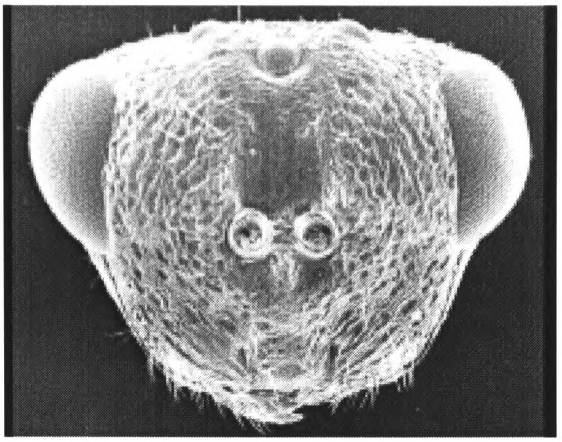


e206

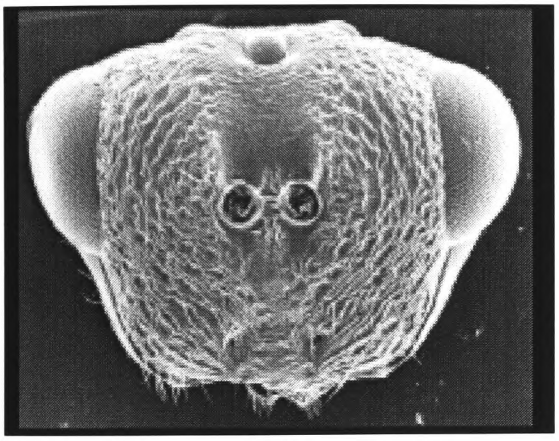
Species 3 : *Tetramesa calamagrostidis*



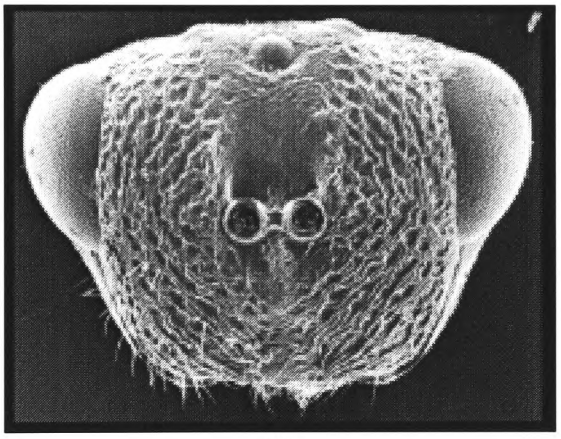
e300



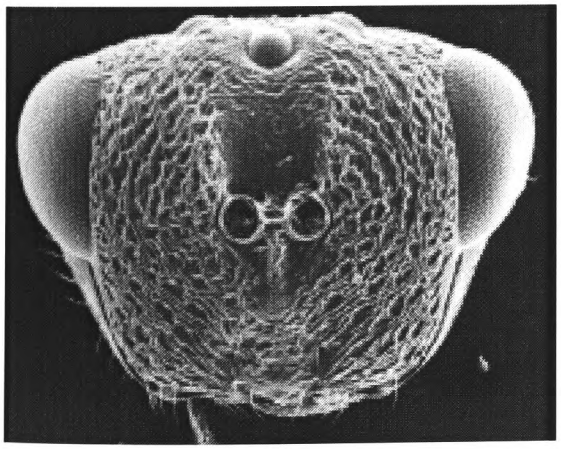
e301



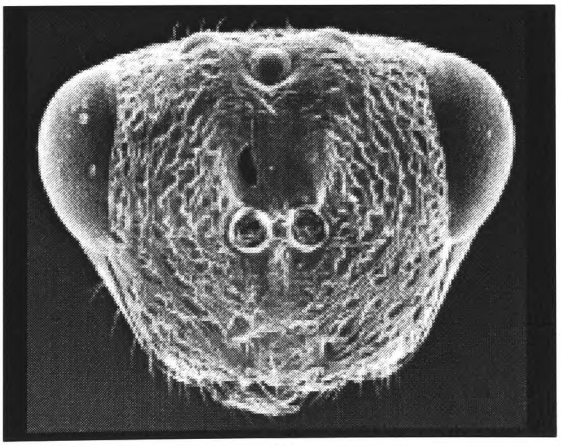
e302



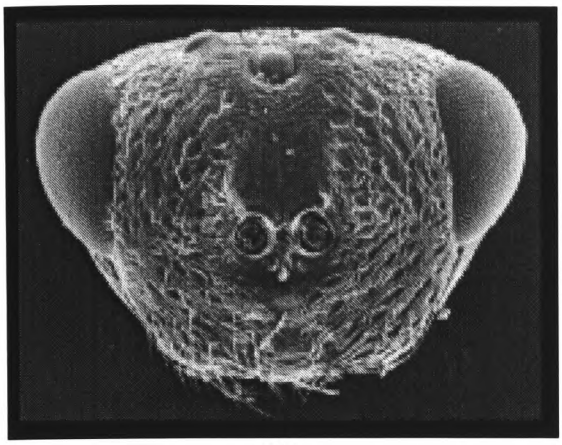
e303



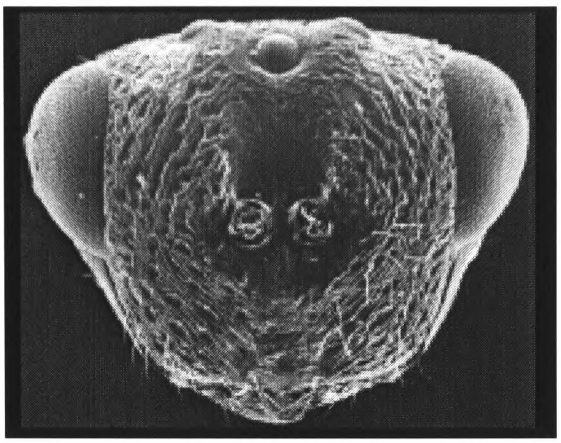
e304



e305

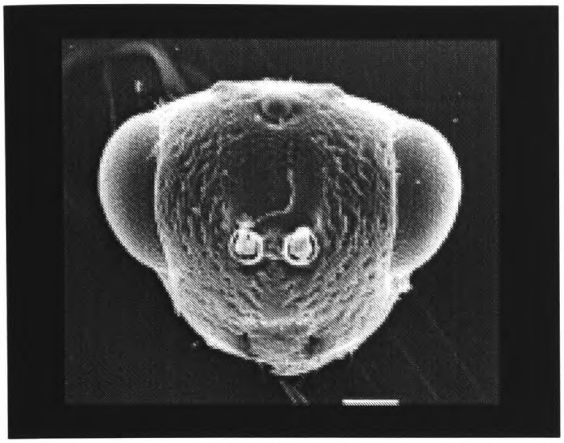


e306

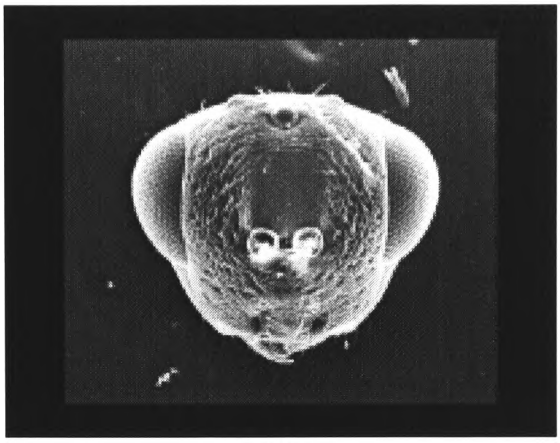


e307

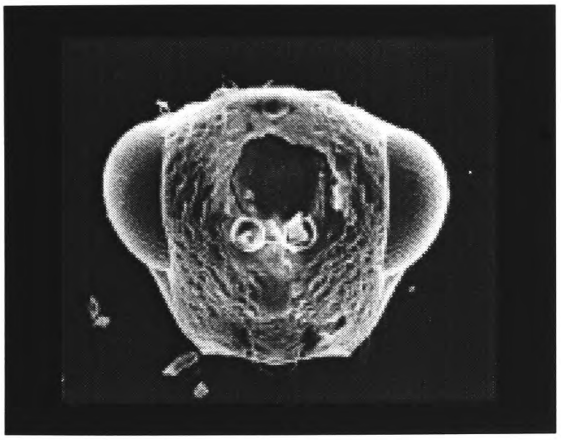
Species 4 : *Tetramesa fulvicollis*



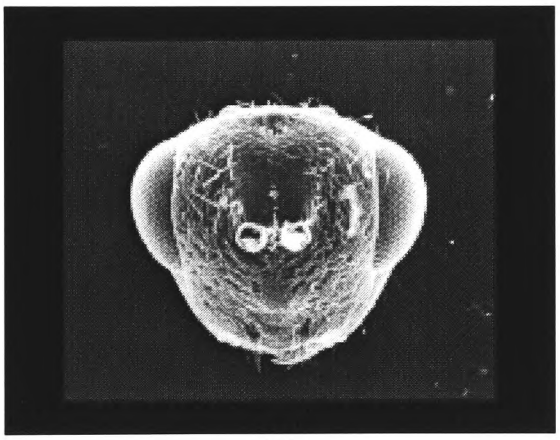
e400



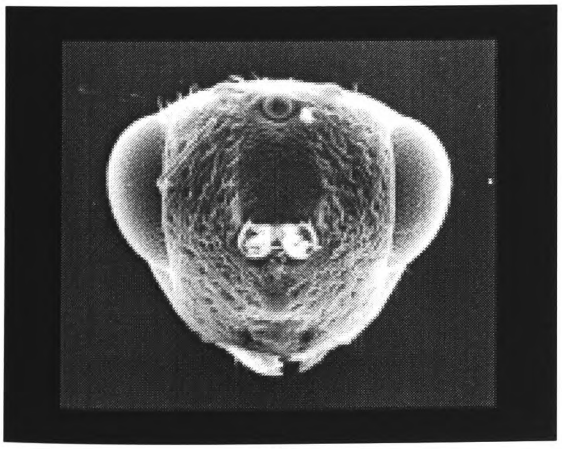
e401



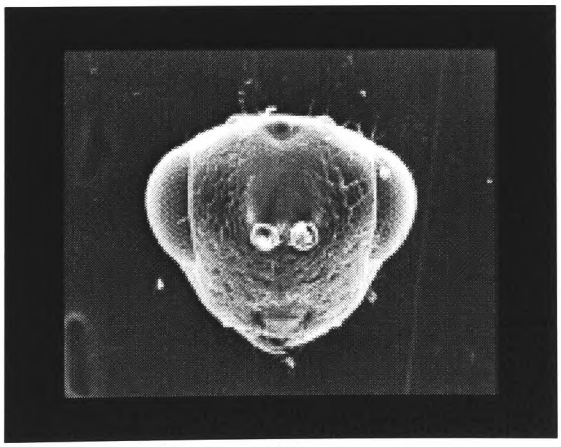
e402



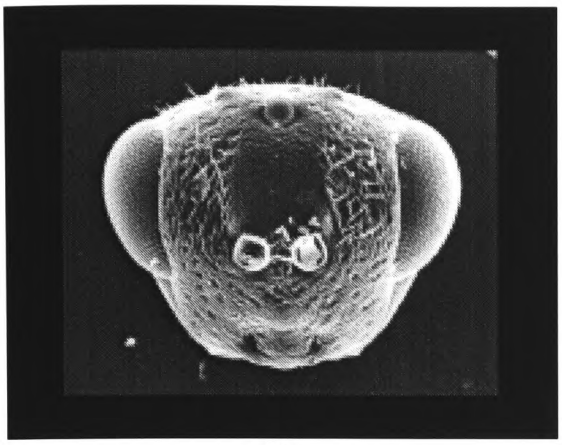
e403



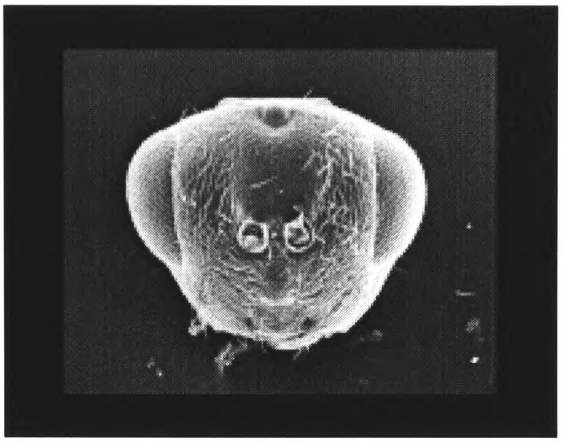
e404



e405

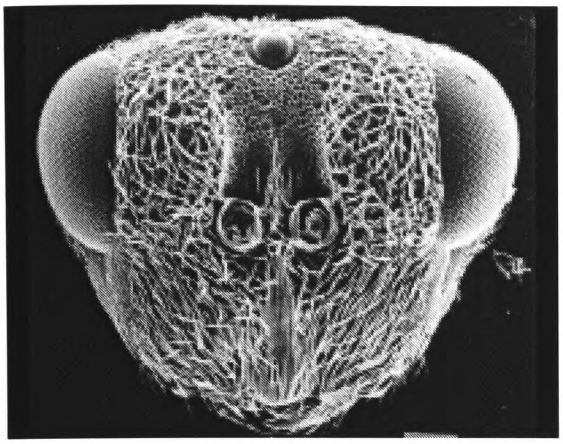


e406

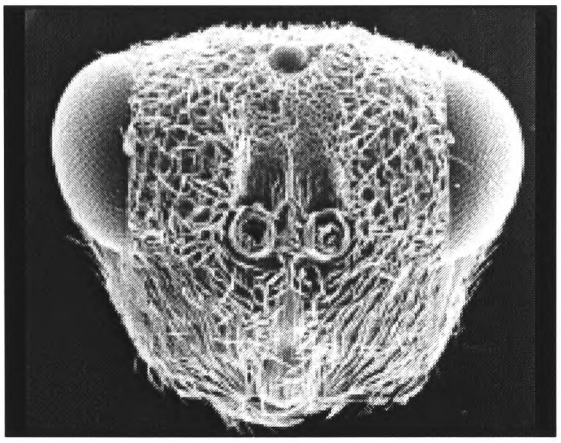


e407

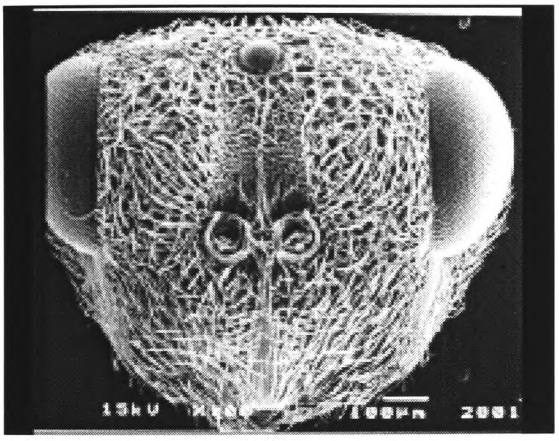
Species 5 : *Tetramesa eximia*



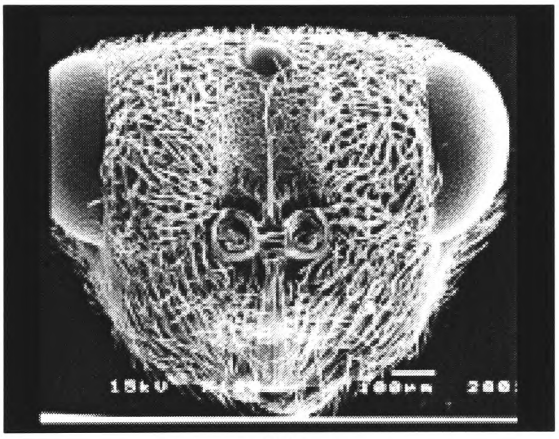
e500



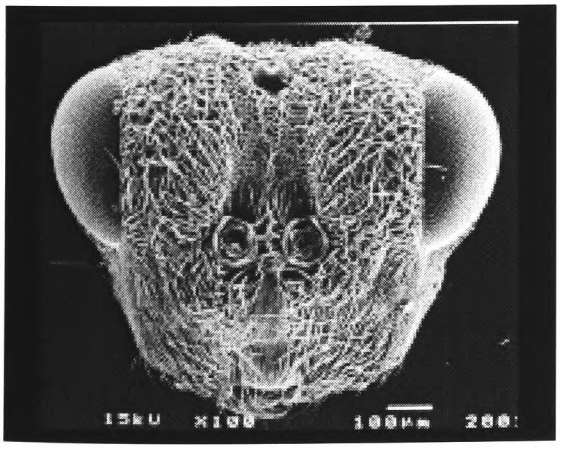
e501



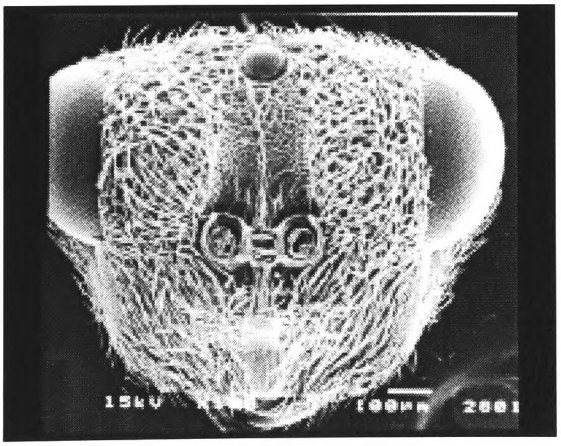
e502



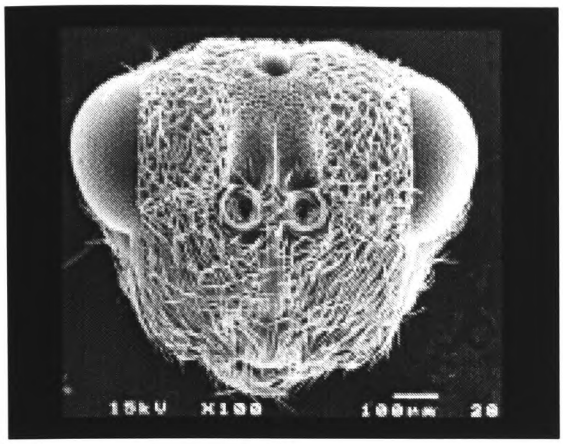
e503



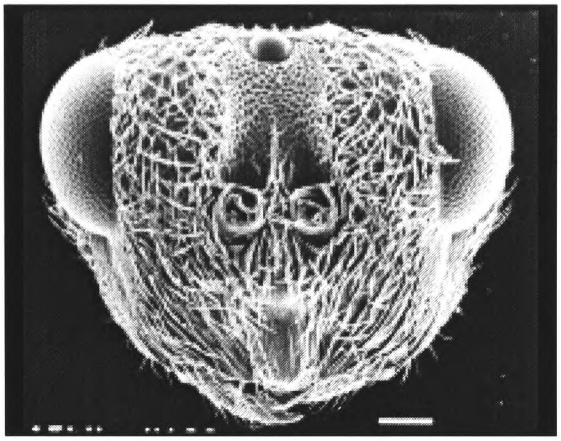
e504



e505



e506

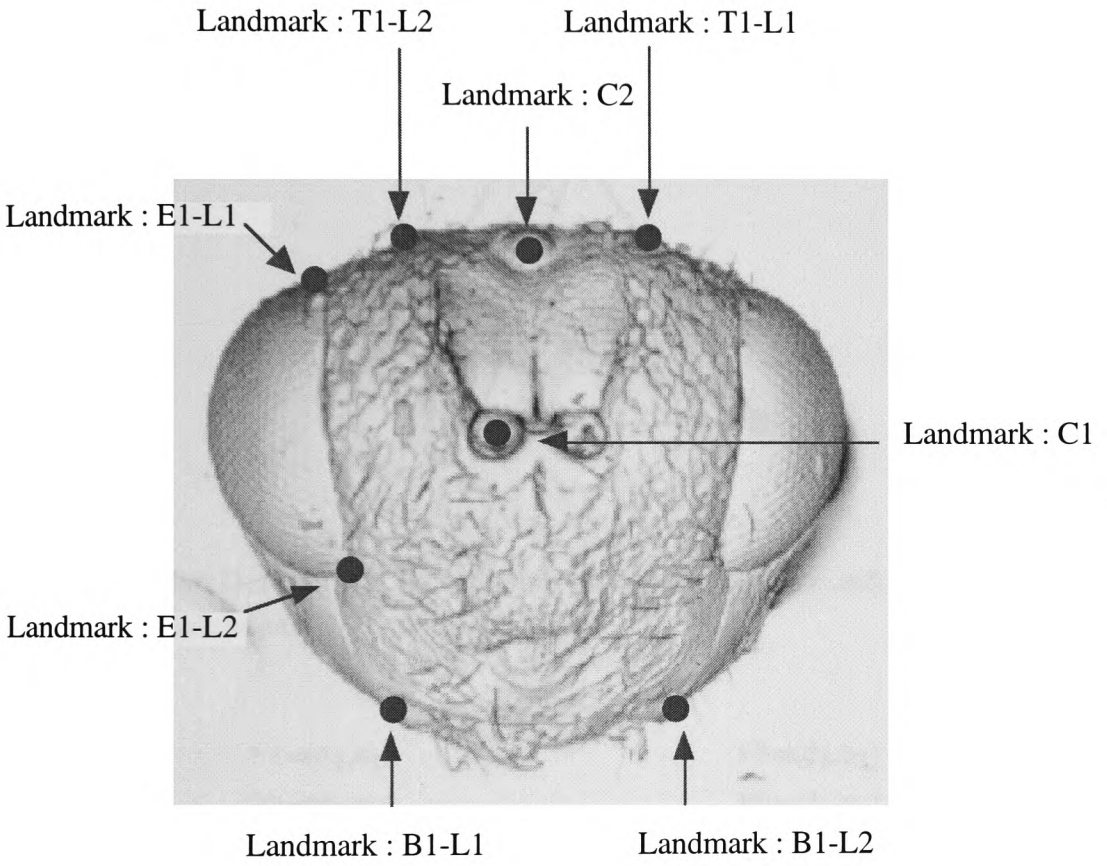


e507

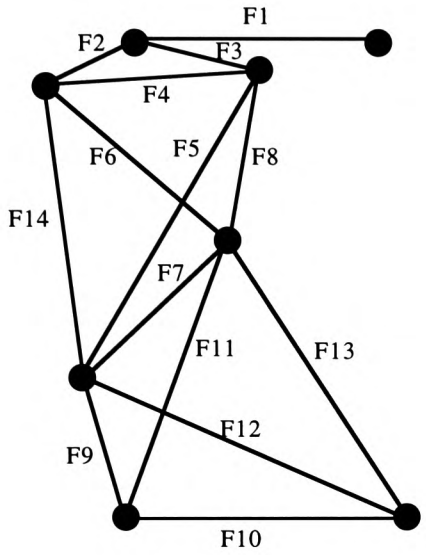
Appendix 2

Taxonomic Features

Taxonomic Landmark Points



Taxomomic Features Derived From Landmark Points



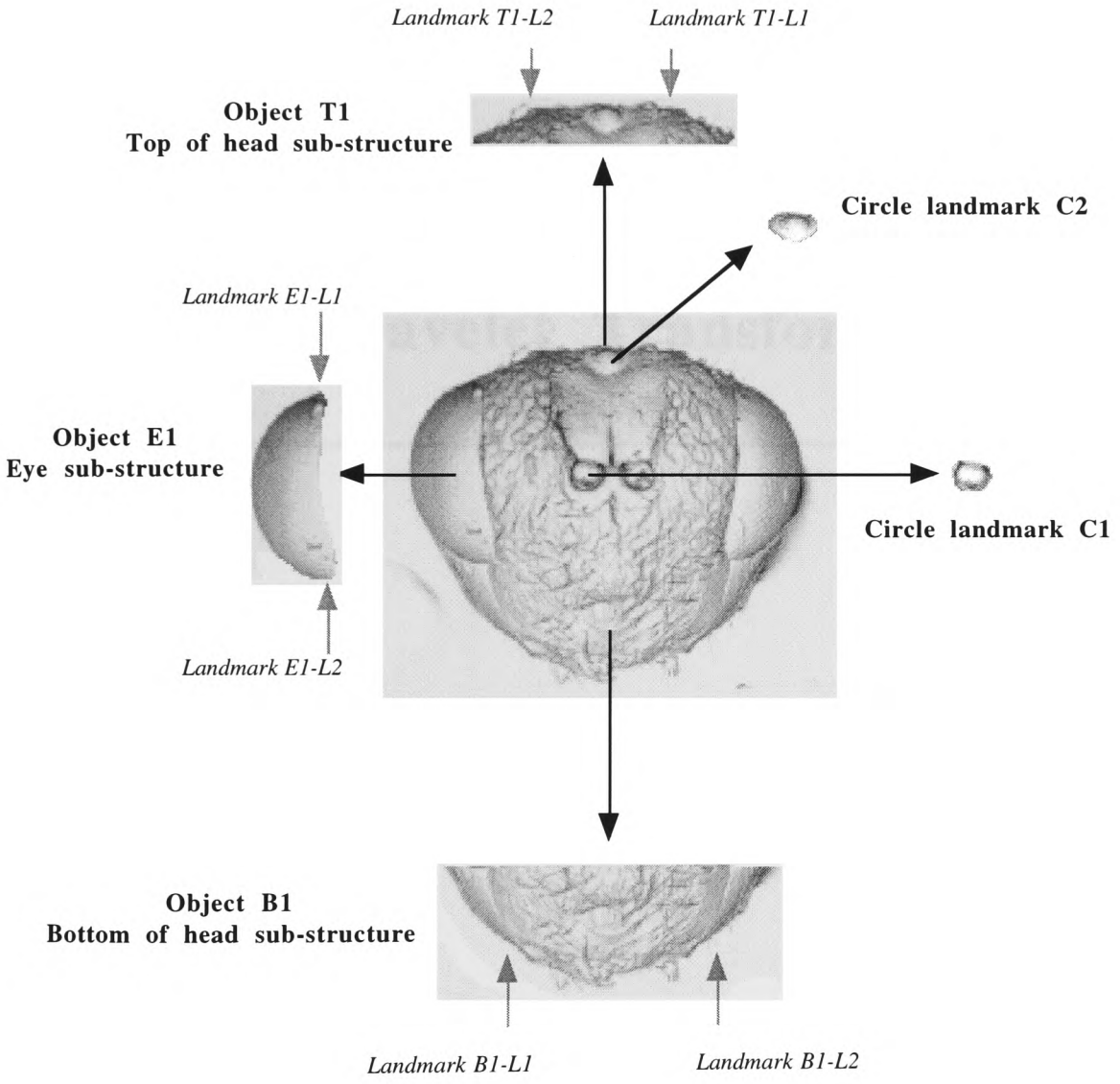
Taxonomic Feature Definitions : Each feature defines the distance and orientation of the two connected landmark points.

$F1=(d_1,\theta_1)$
$F2=(d_2,\theta_2)$
$F3=(d_3,\theta_3)$
$F4=(d_4,\theta_4)$
$F5=(d_5,\theta_5)$
$F6=(d_6,\theta_6)$
$F7=(d_7,\theta_7)$
$F8=(d_8,\theta_8)$
$F9=(d_9,\theta_9)$
$F10=(d_{10},\theta_{10})$
$F11=(d_{11},\theta_{11})$
$F12=(d_{12},\theta_{12})$
$F13=(d_{13},\theta_{13})$
$F14=(d_{14},\theta_{14})$

Appendix 3

Wasp Head Sub-structures

Sub-structures and Associated Landmark Points

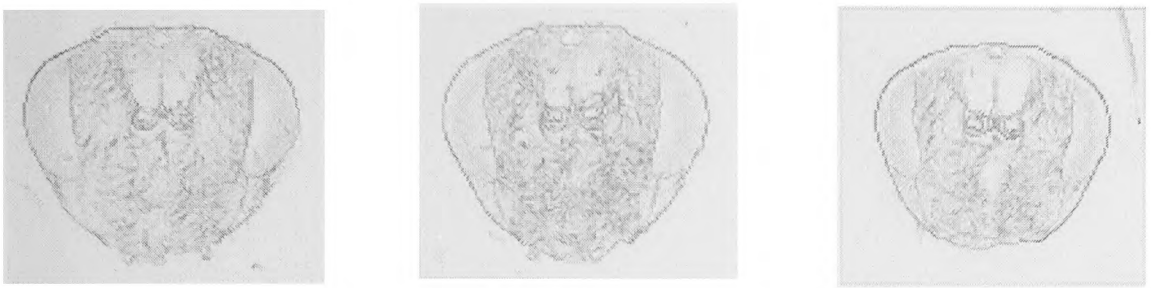


Appendix 4

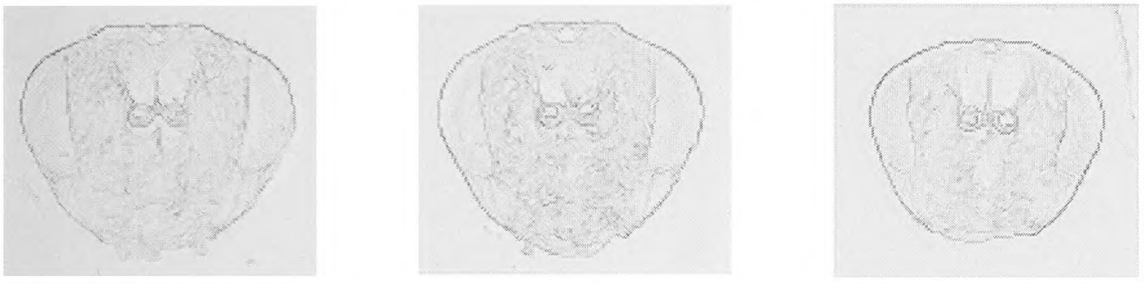
Wavelet Transform Images

Species 1 : *Tetramesa linearis*

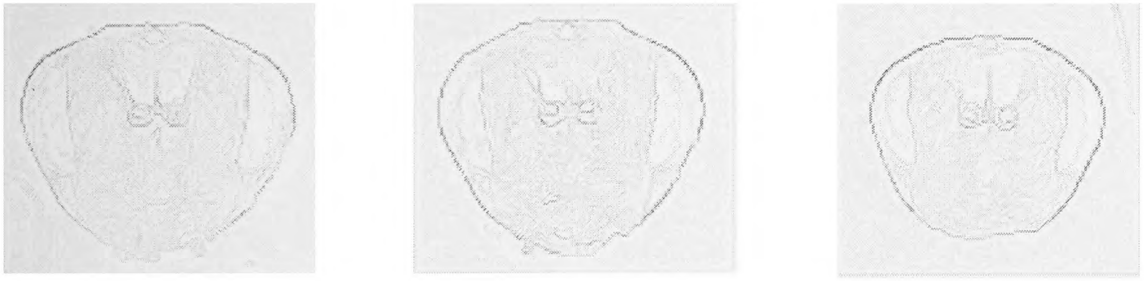
e100 e101 e102



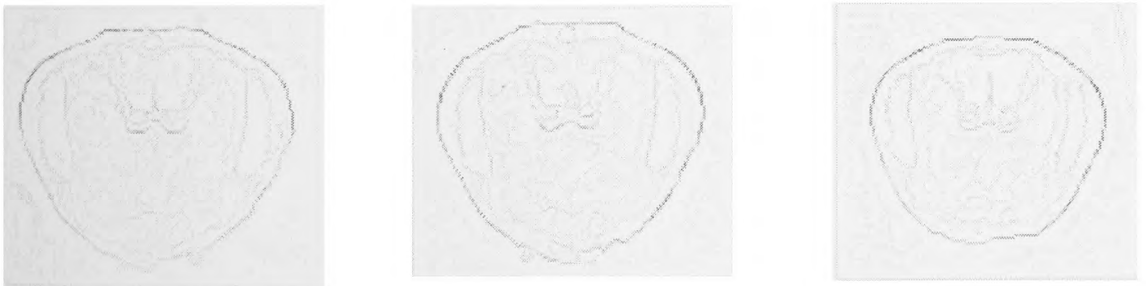
$\sigma=2^0$



$\sigma=2^1$

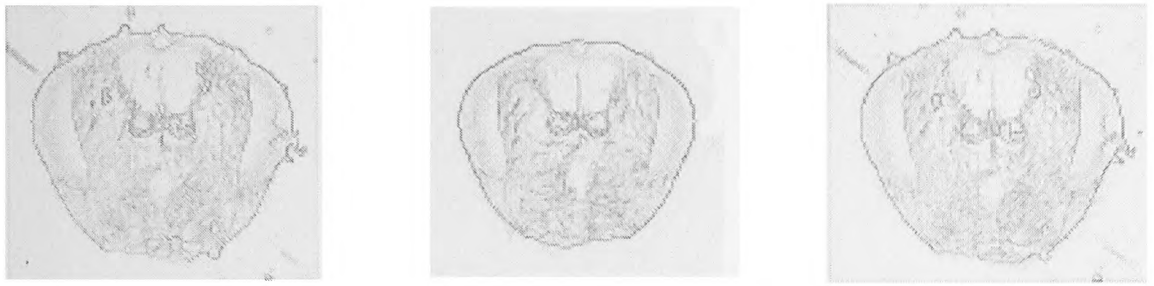


$\sigma=2^2$

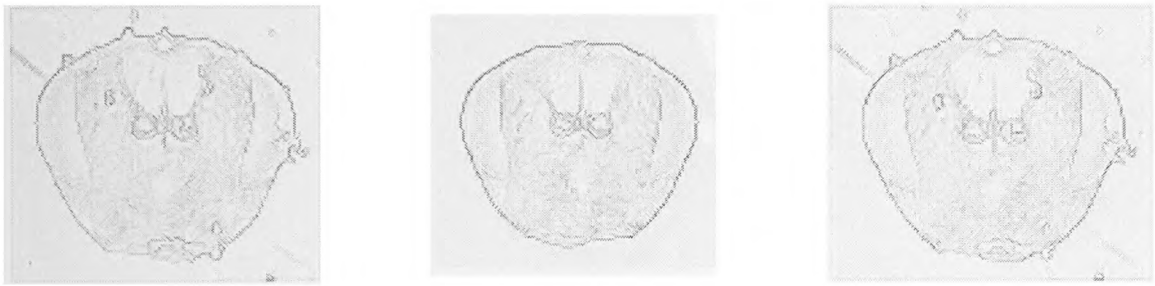


$\sigma=2^3$

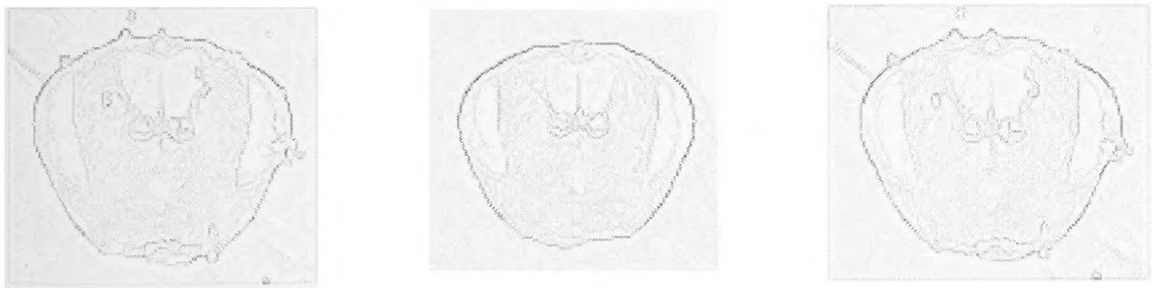
e103 e104 e105



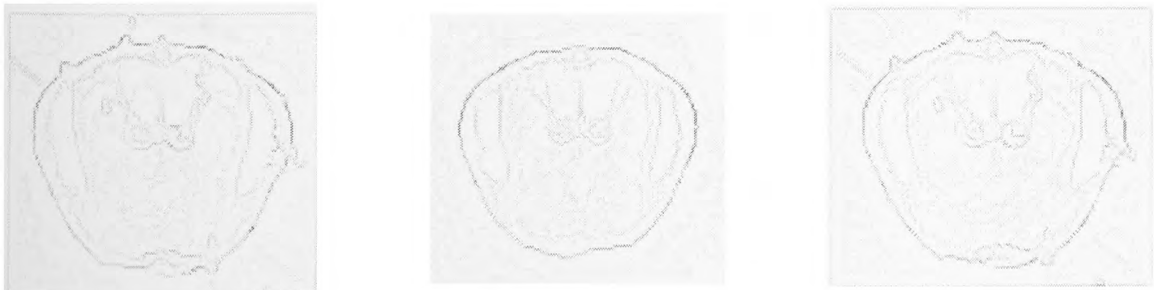
$\sigma=2^0$



$\sigma=2^1$

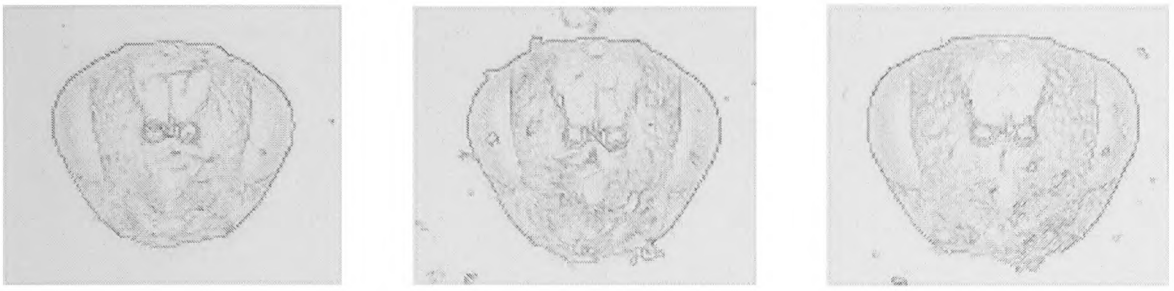


$\sigma=2^2$

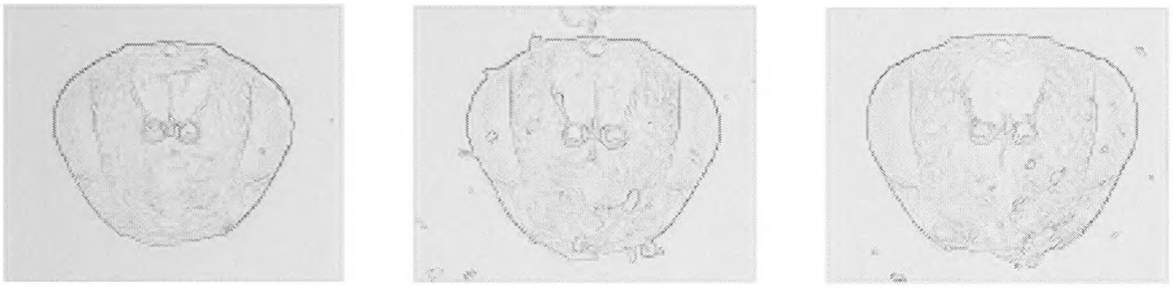


$\sigma=2^3$

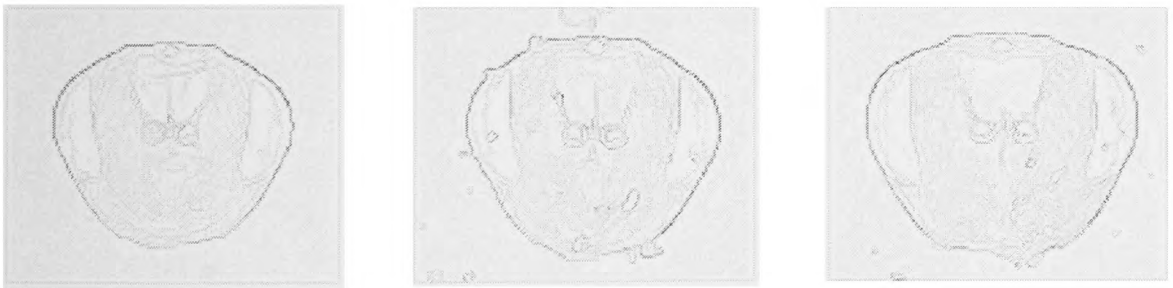
e106 e107 e108



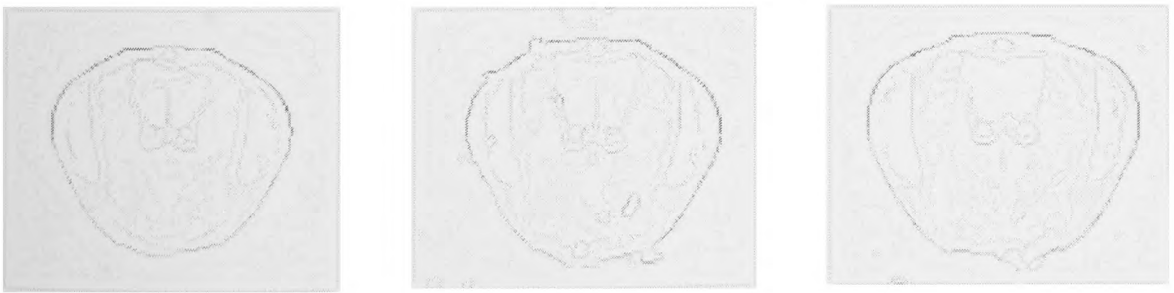
$\sigma=2^0$



$\sigma=2^1$

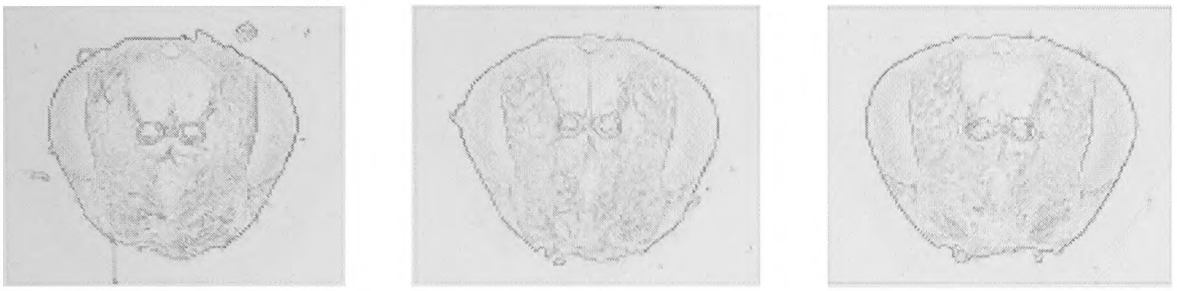


$\sigma=2^2$

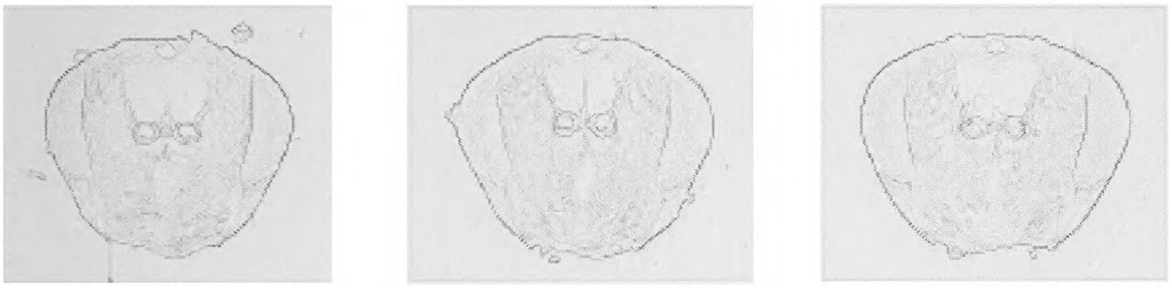


$\sigma=2^3$

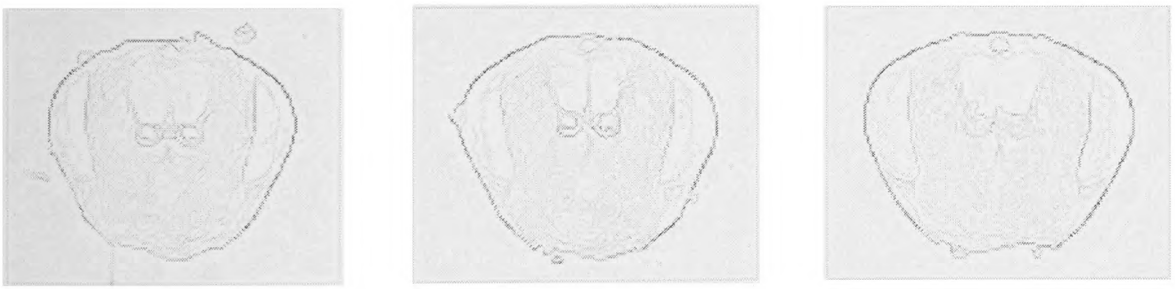
e109 e110 e111



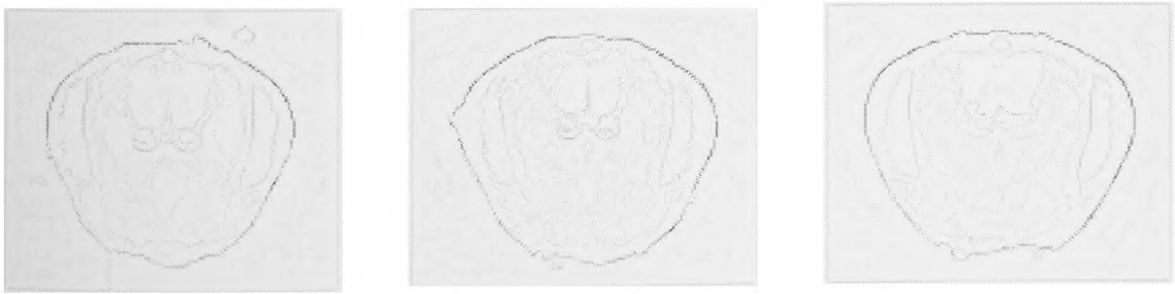
$\sigma=2^0$



$\sigma=2^1$



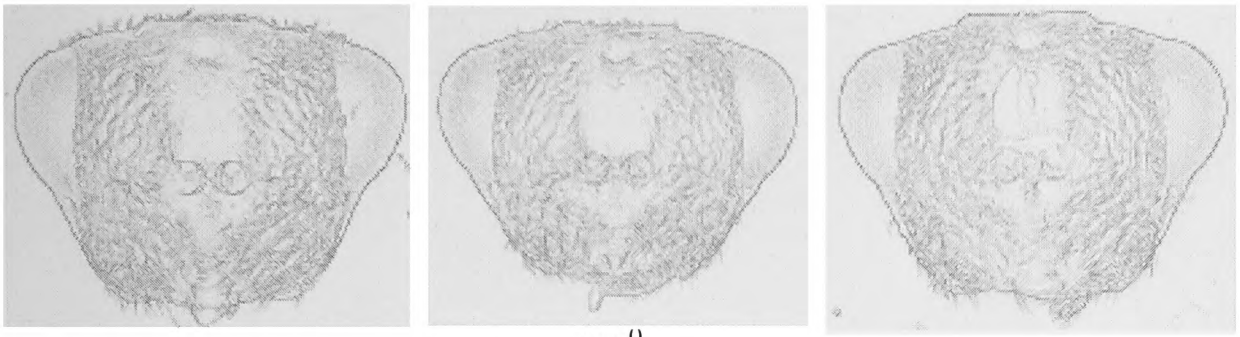
$\sigma=2^2$



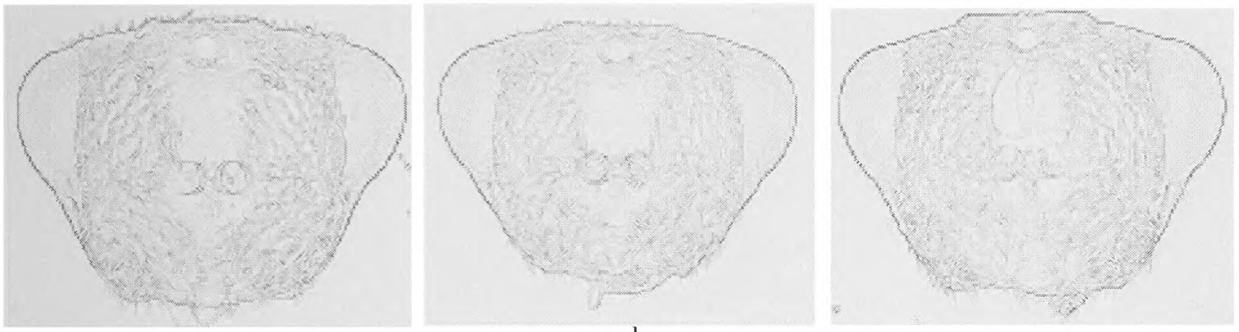
$\sigma=2^3$

Species 2 : *Tetramesa hyalipennis*

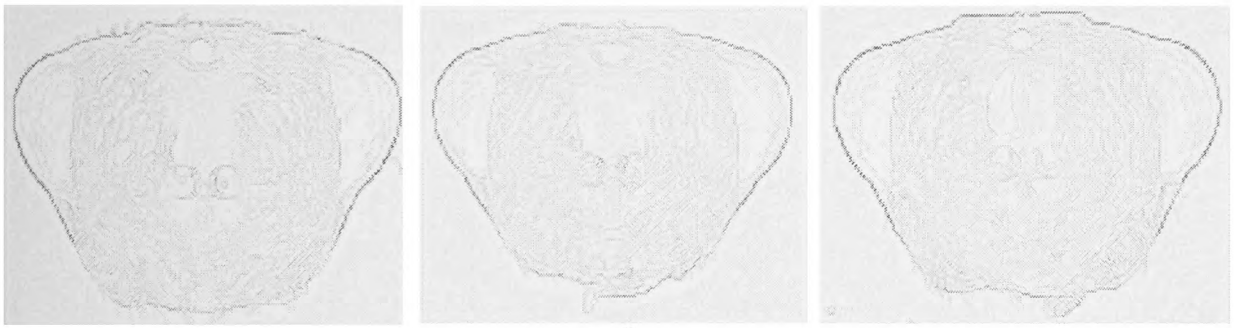
e200 e201 e202



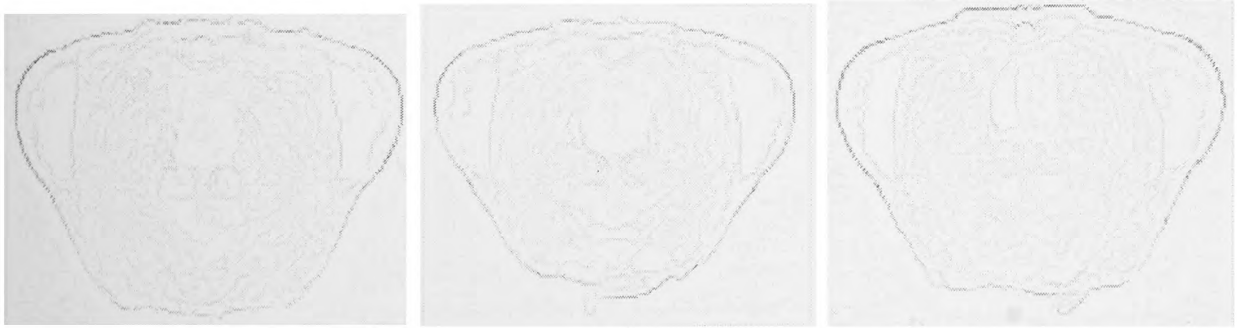
$\sigma=2^0$



$\sigma=2^1$

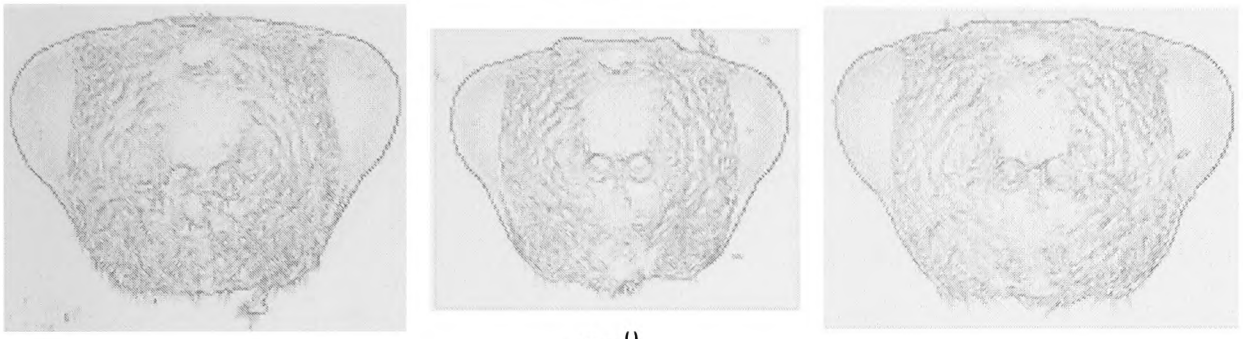


$\sigma=2^2$

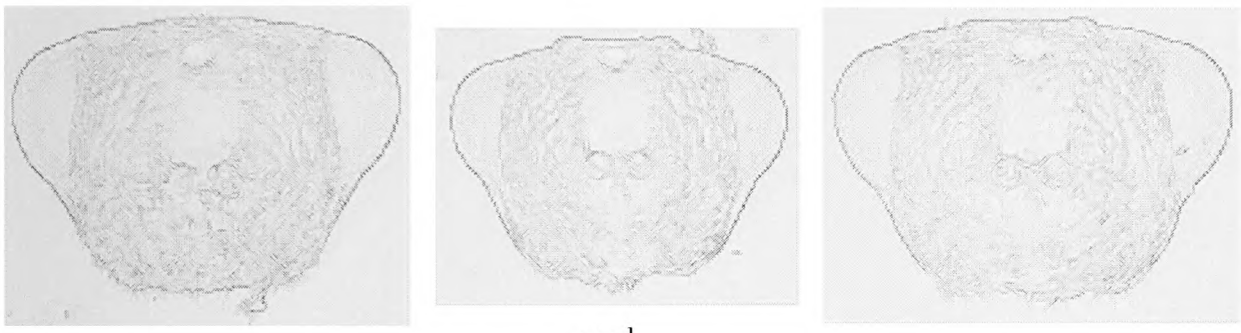


$\sigma=2^3$

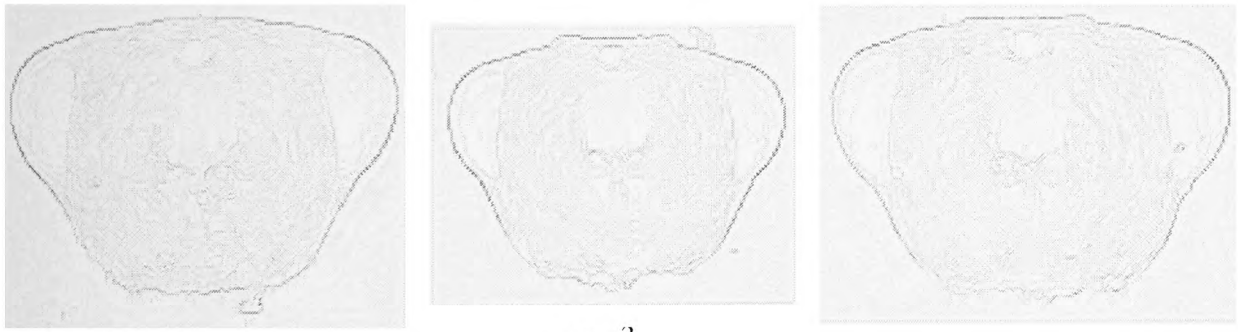
e203 e204 e205



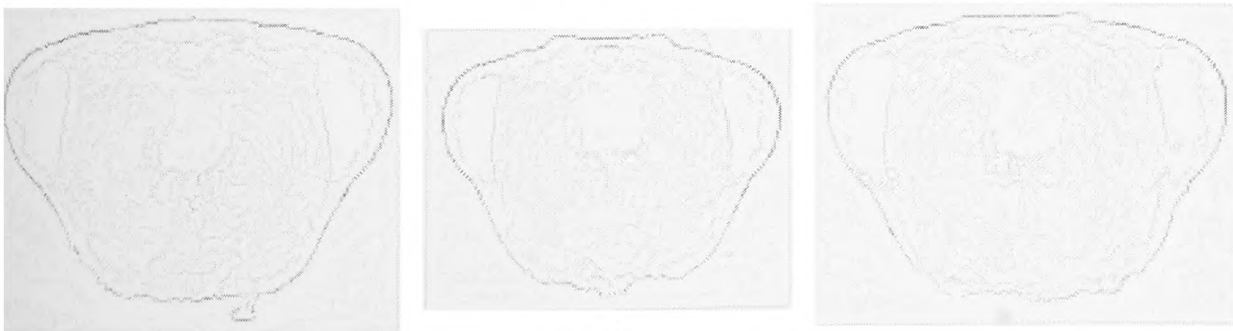
$\sigma=2^0$



$\sigma=2^1$

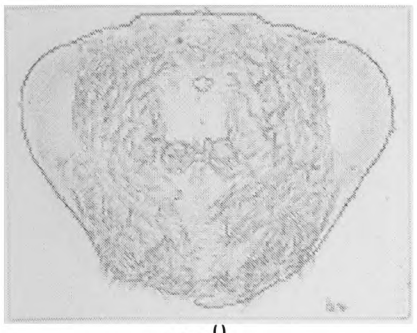


$\sigma=2^2$

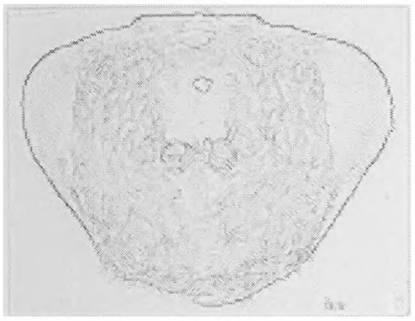


$\sigma=2^3$

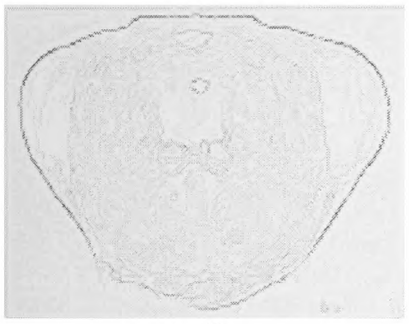
e206



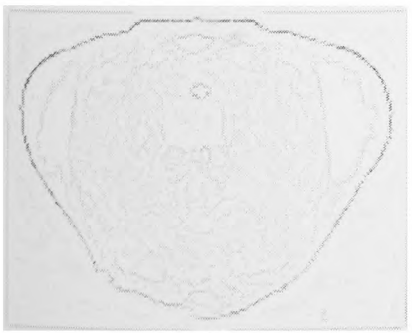
$\sigma=2^0$



$\sigma=2^1$

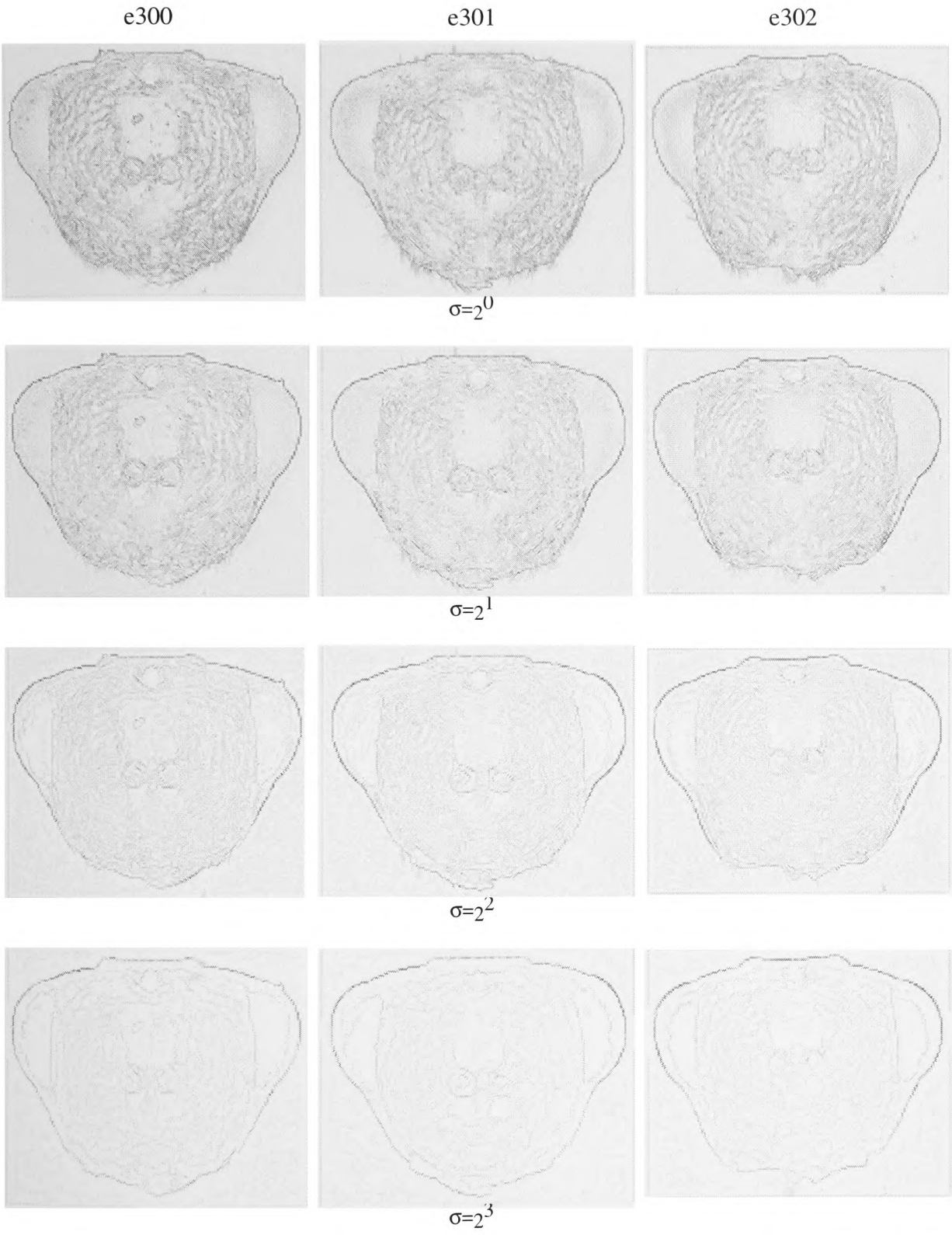


$\sigma=2^2$

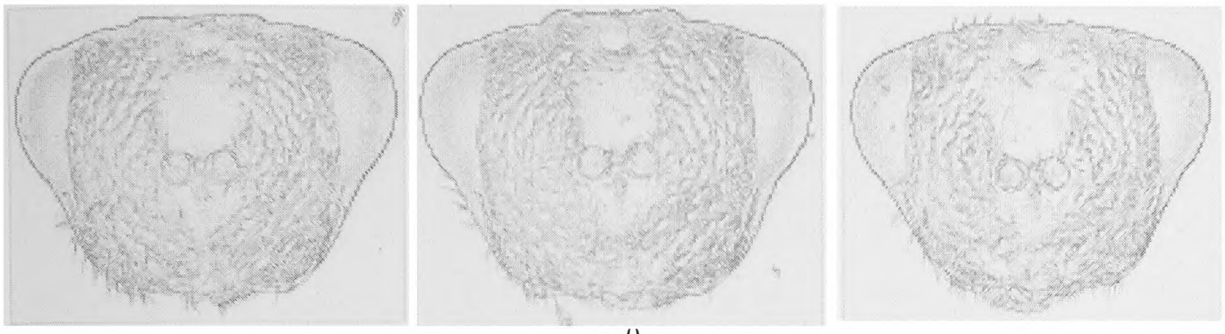


$\sigma=2^3$

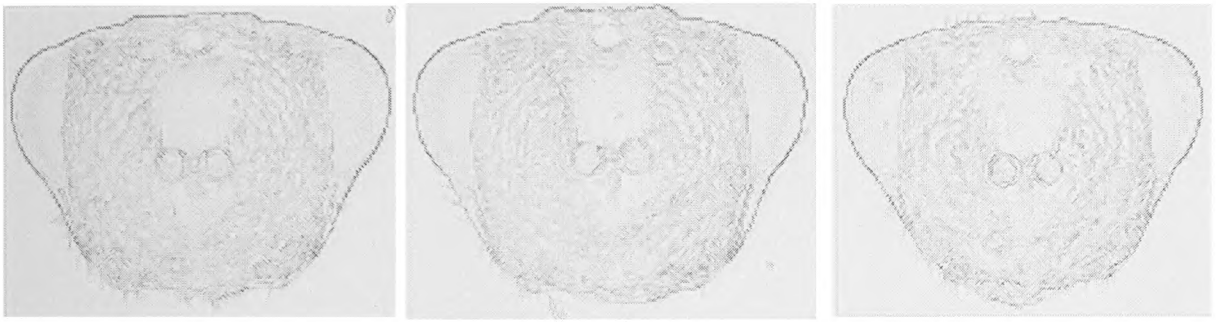
Species 3 : *Tetramesa calamagrostidis*



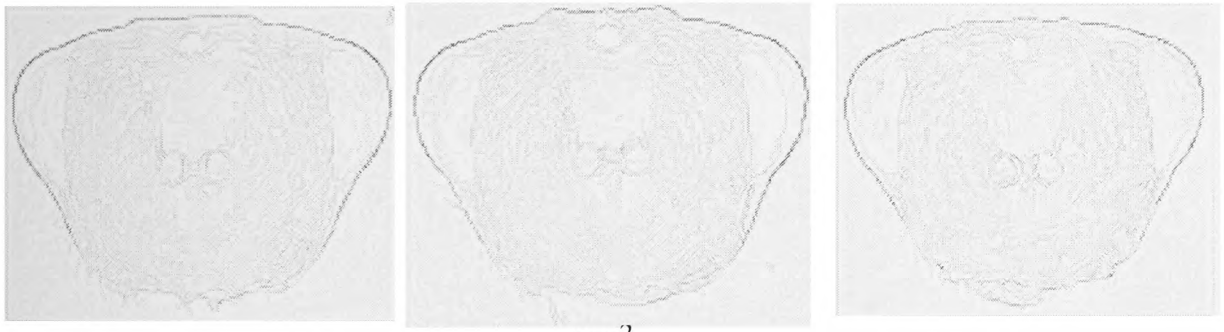
e303 e304 e305



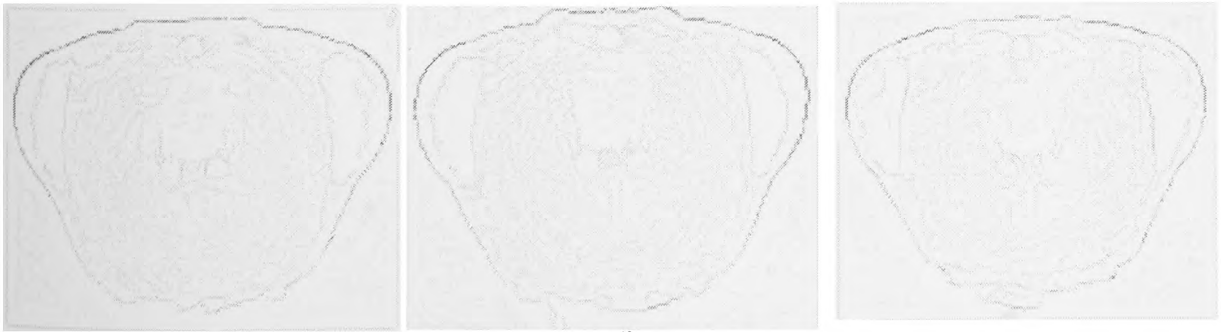
$\sigma=2^0$



$\sigma=2^1$

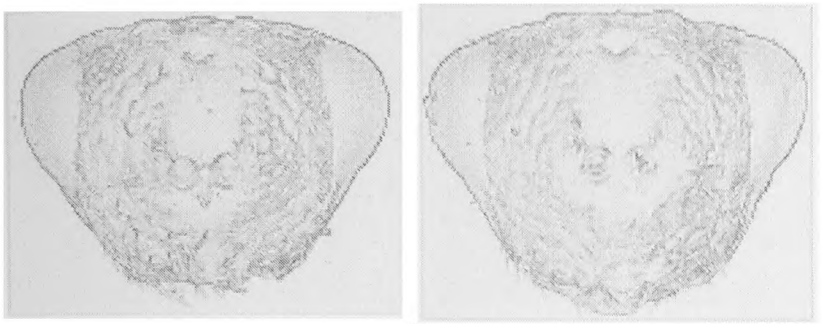


$\sigma=2^2$

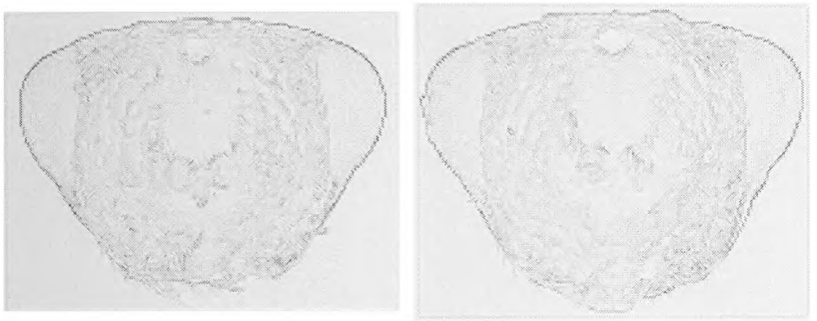


$\sigma=2^3$

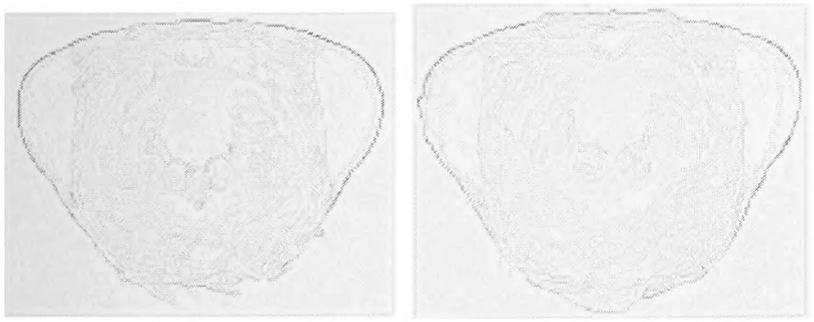
e306 e307



$\sigma=2^0$



$\sigma=2^1$

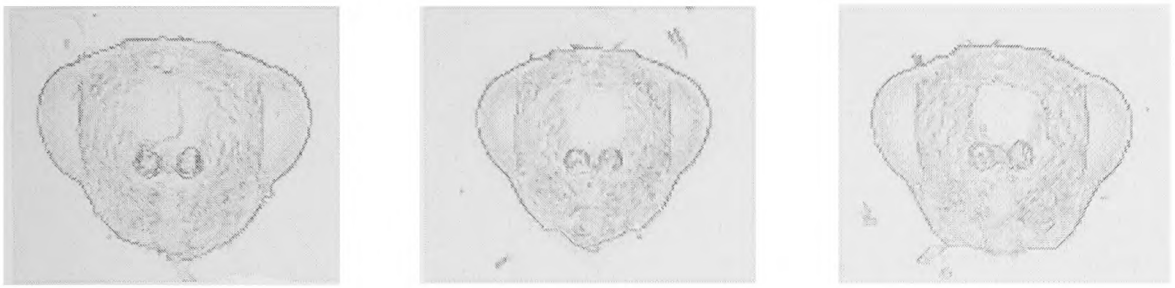


$\sigma=2^2$

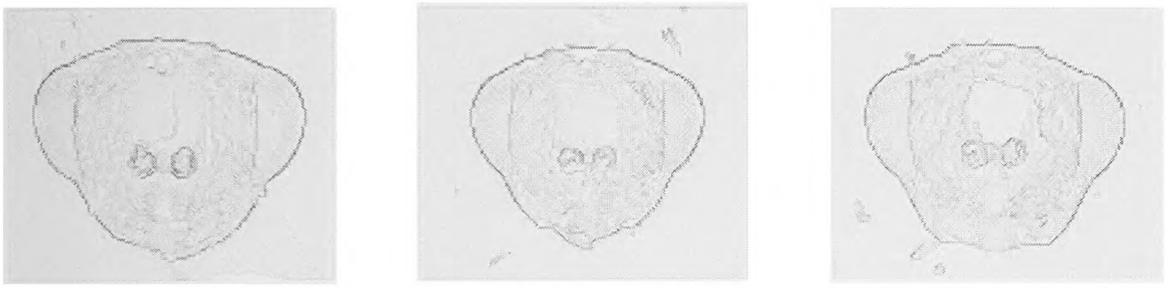
$\sigma=2^3$

Species 4 : *Tetramesa fulvicollis*

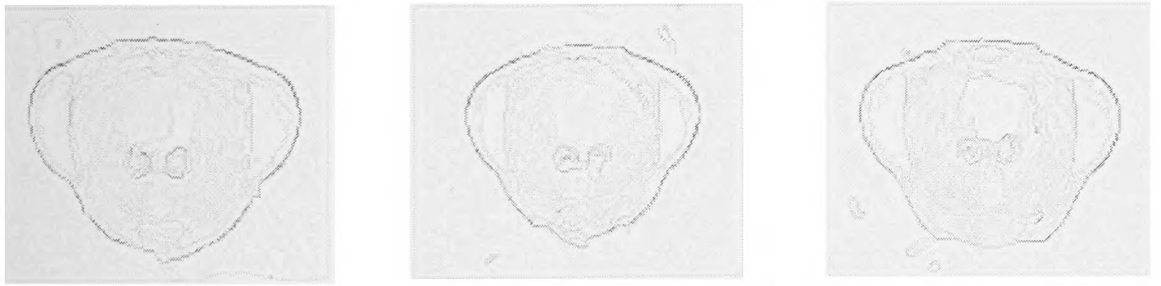
e400 e401 e402



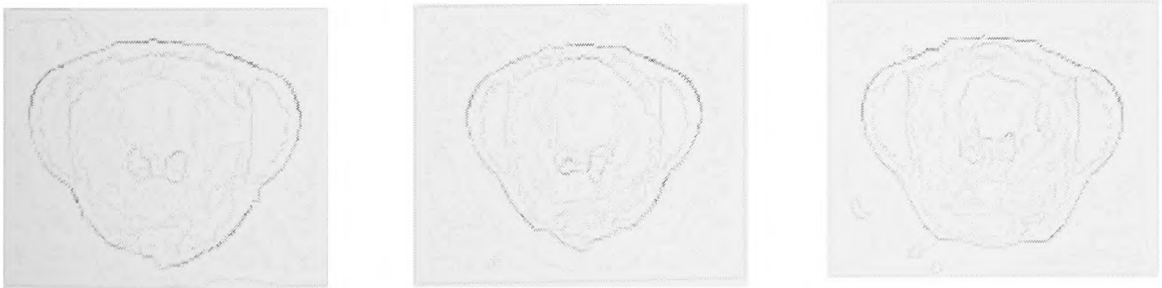
$\sigma=2^0$



$\sigma=2^1$

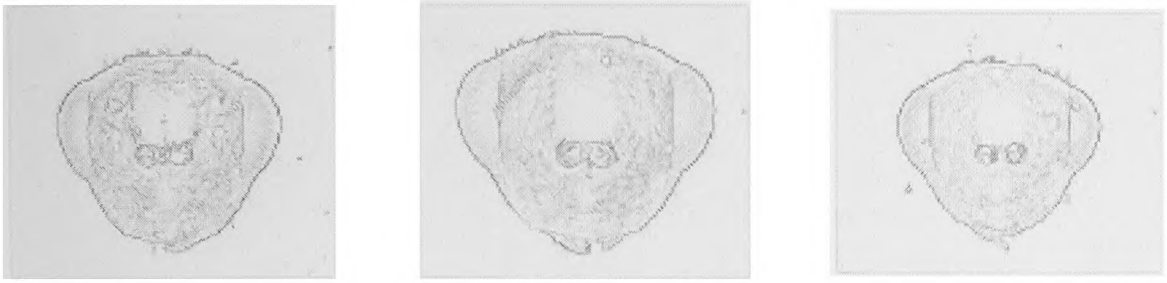


$\sigma=2^2$

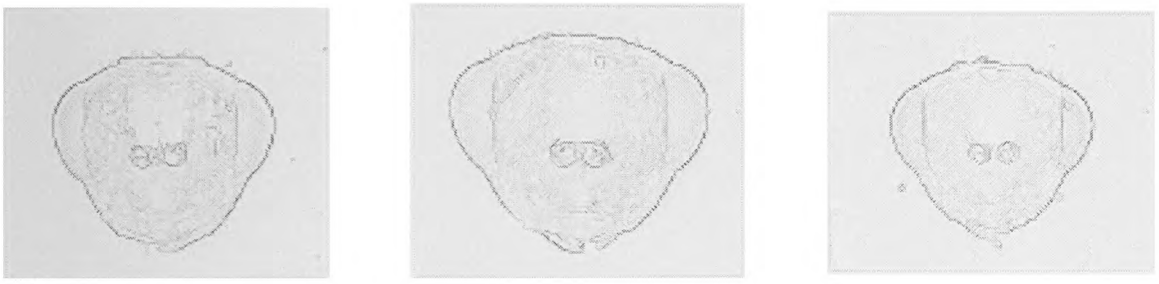


$\sigma=2^3$

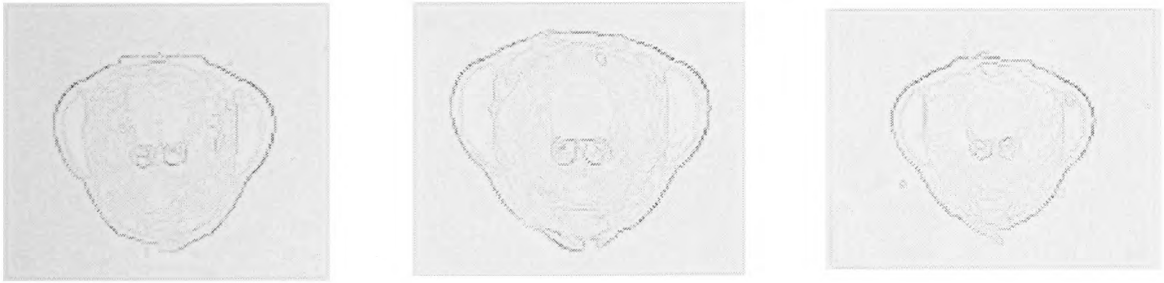
e403 e404 e405



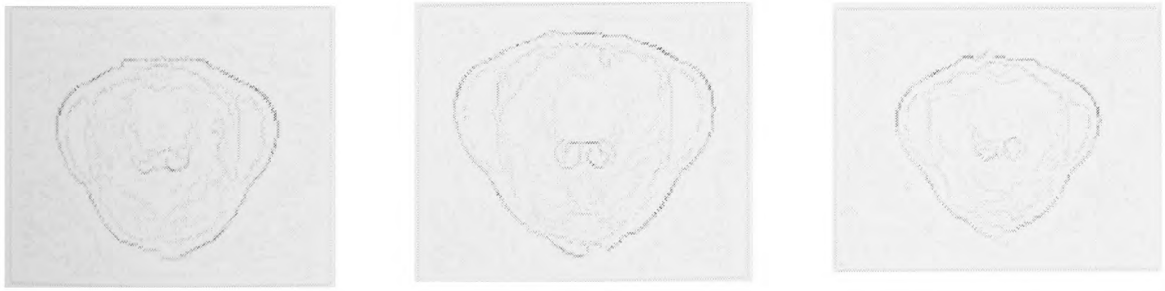
$\sigma=2^0$



$\sigma=2^1$

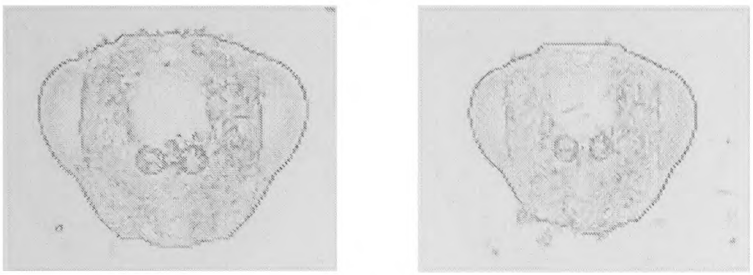


$\sigma=2^2$

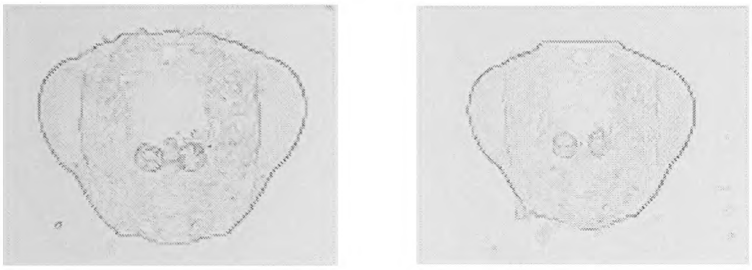


$\sigma=2^3$

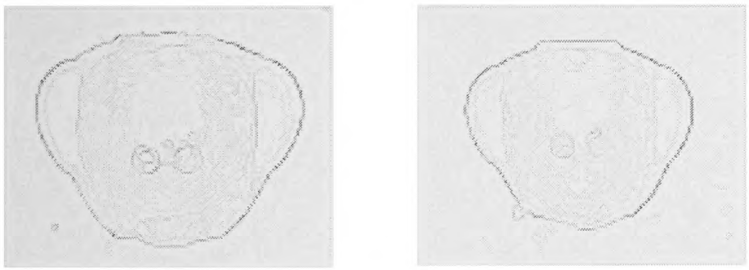
e406 e407



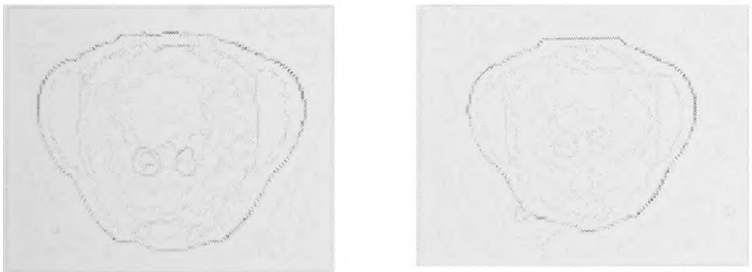
$\sigma=2^0$



$\sigma=2^1$

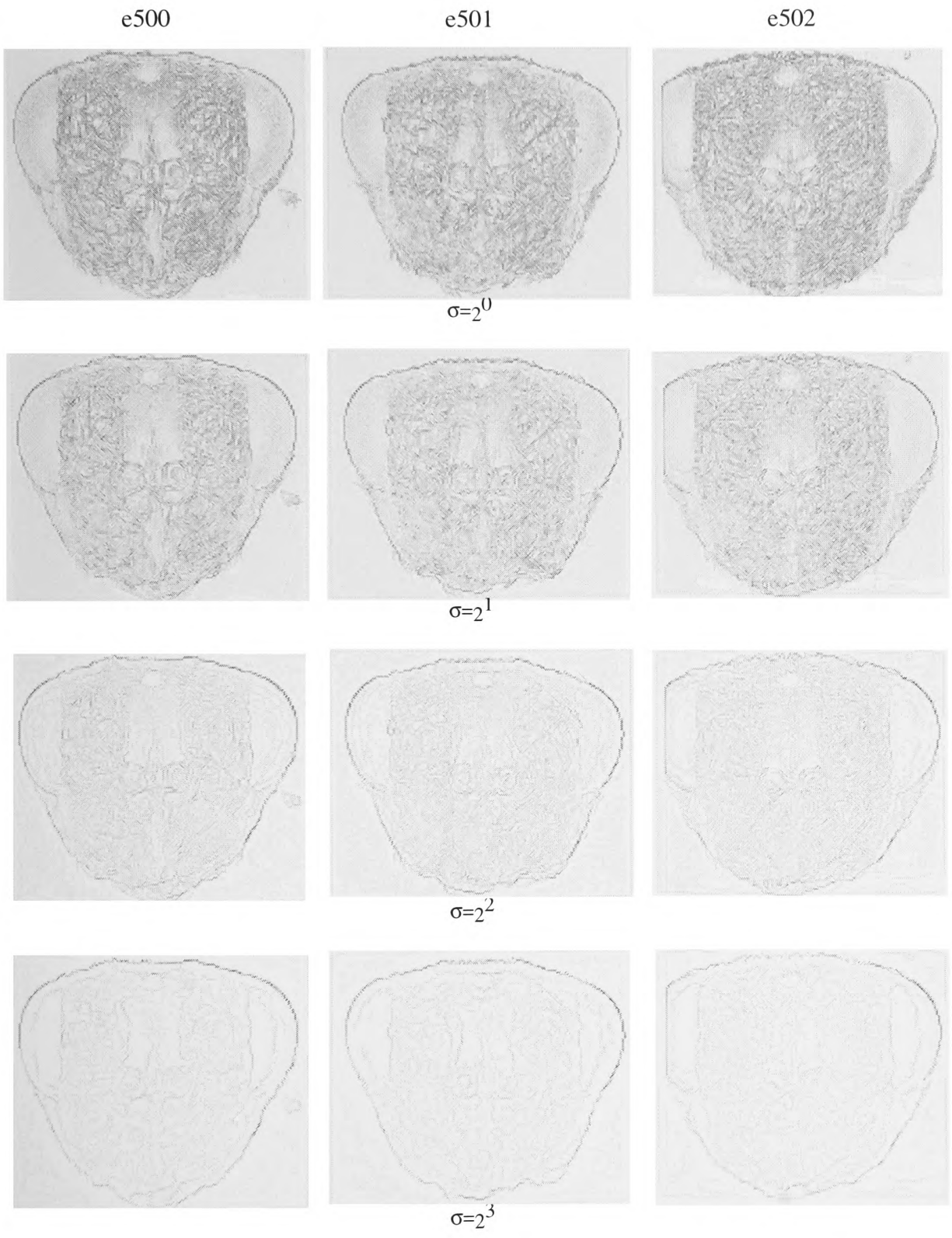


$\sigma=2^2$

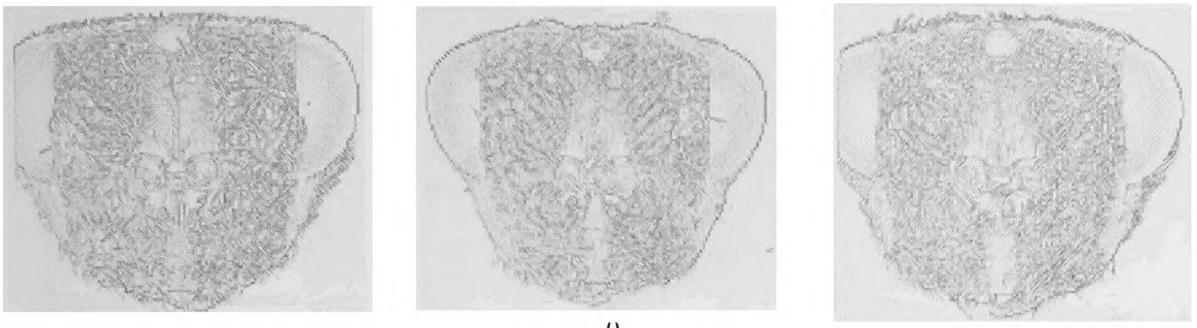


$\sigma=2^3$

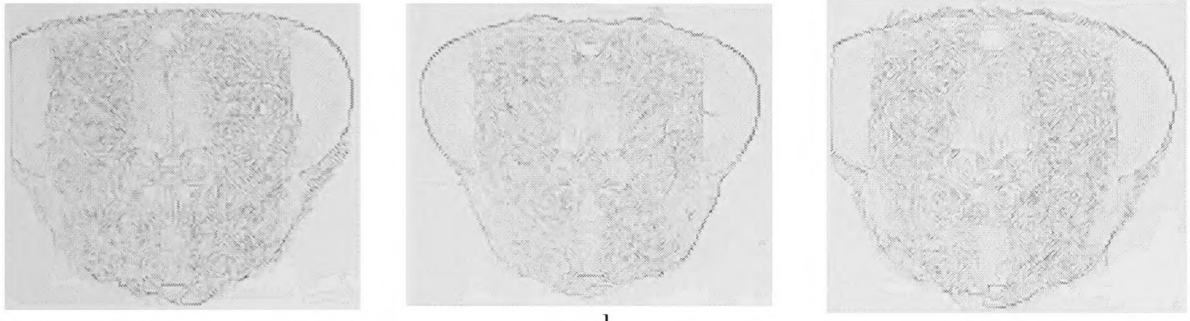
Species 5 : *Tetramesa eximia*



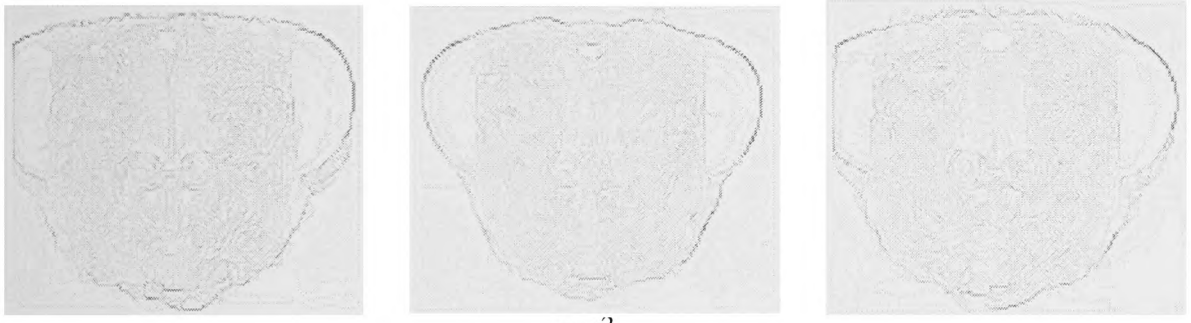
e503 e504 e505



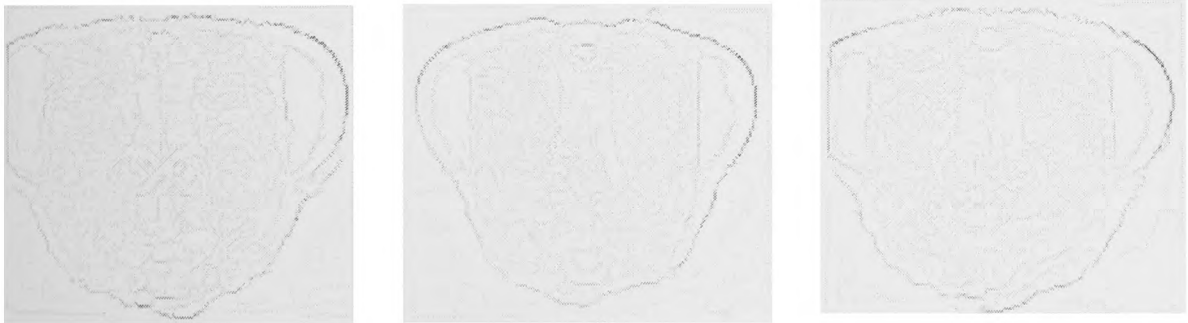
$\sigma=2^0$



$\sigma=2^1$

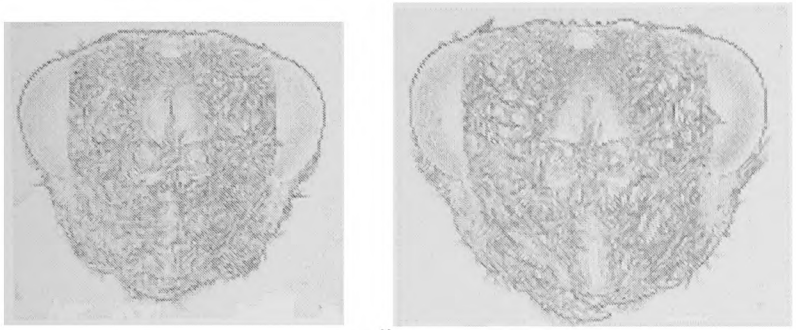


$\sigma=2^2$

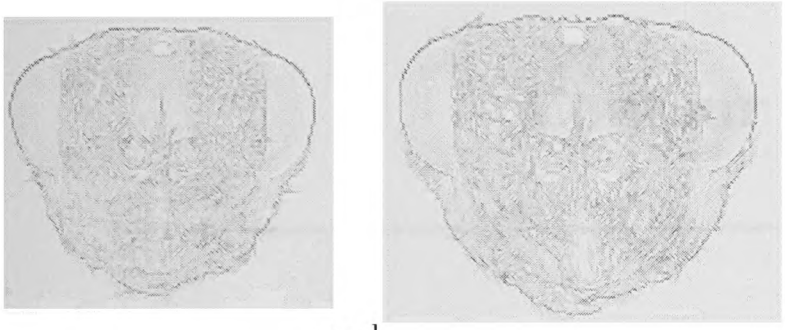


$\sigma=2^3$

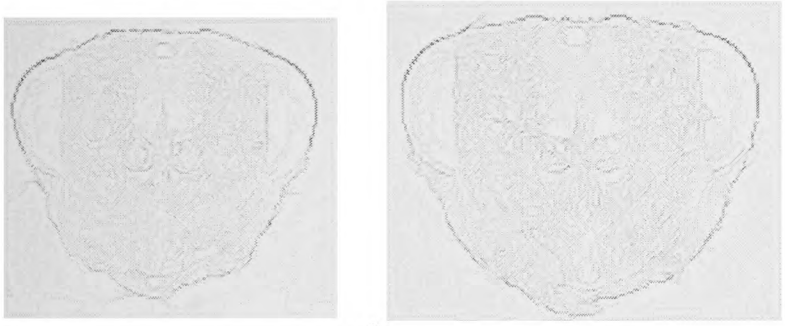
e506 e507



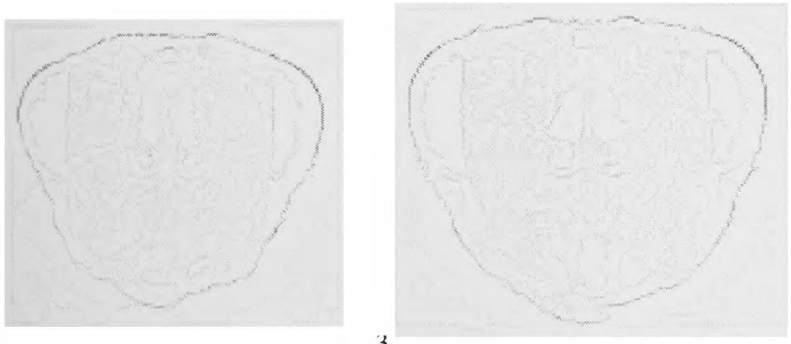
$\sigma=2^0$



$\sigma=2^1$



$\sigma=2^2$

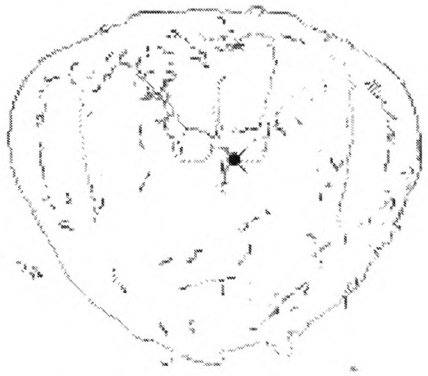


$\sigma=2^3$

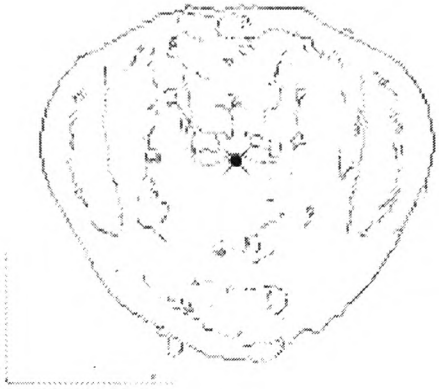
Appendix 5

Contour Map Images

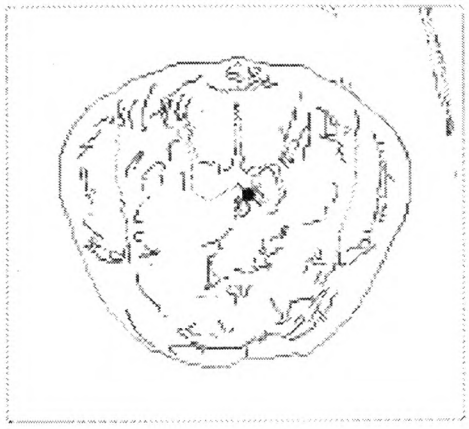
Species 1 : *Tetramesa linearis*



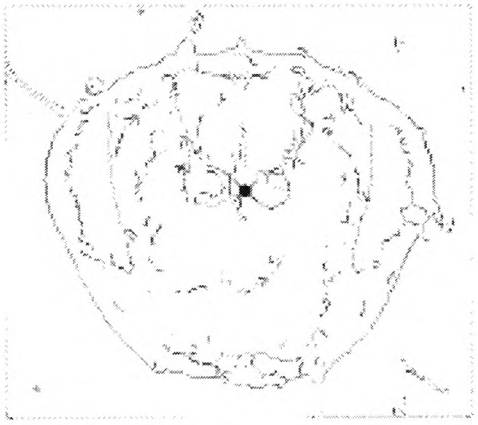
e100



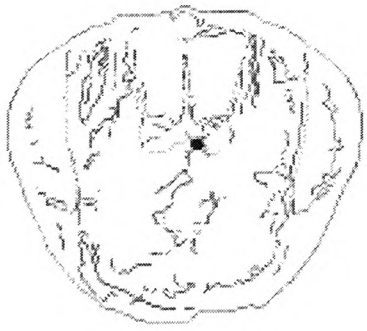
e101



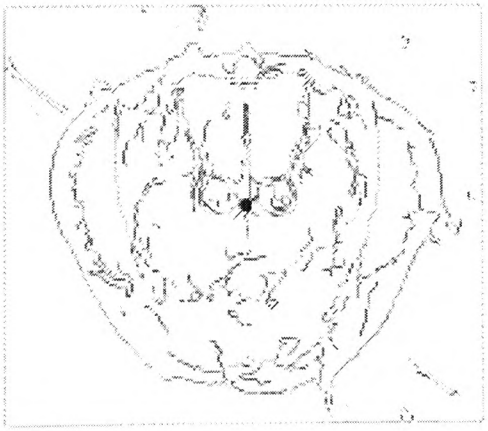
e102



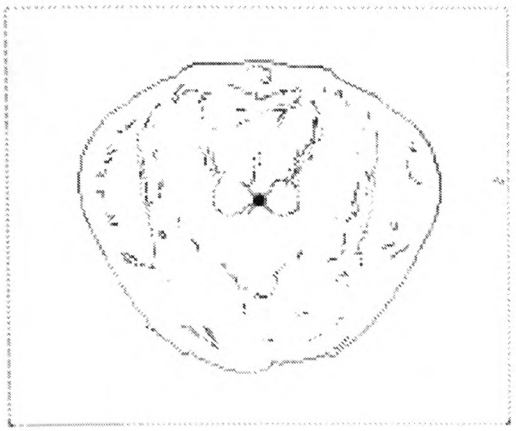
e103



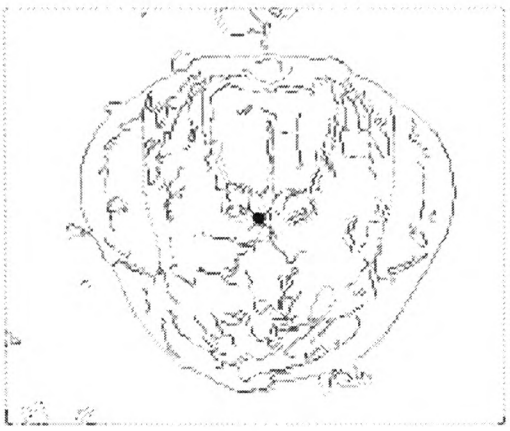
e104



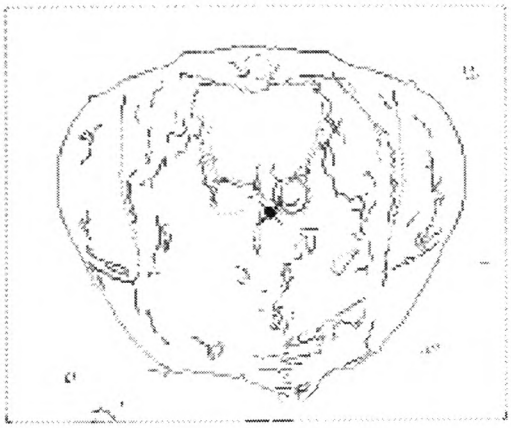
e105



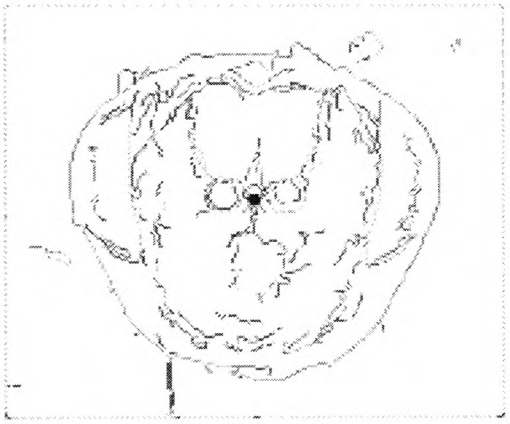
e106



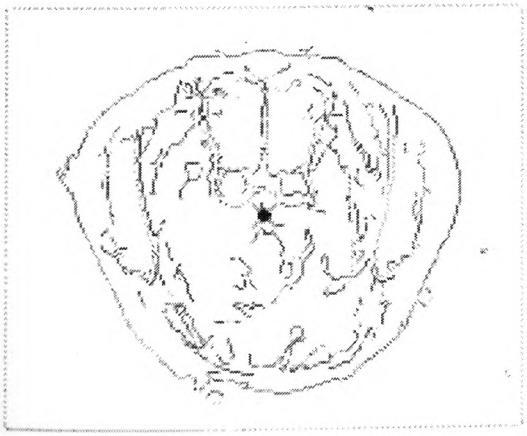
e107



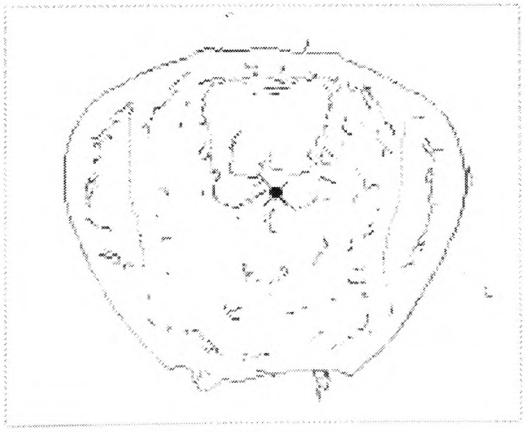
e108



e109

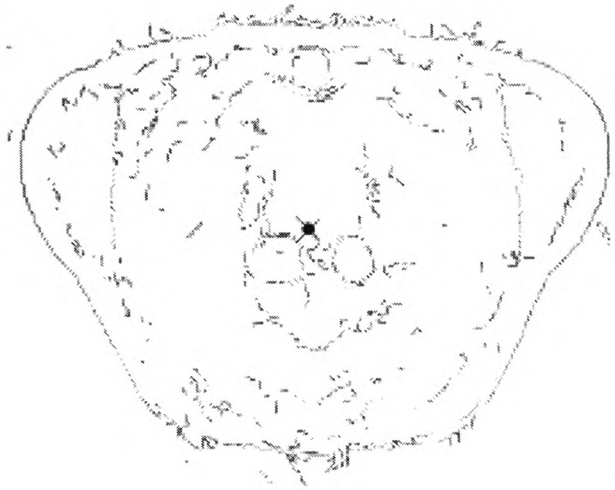


e110

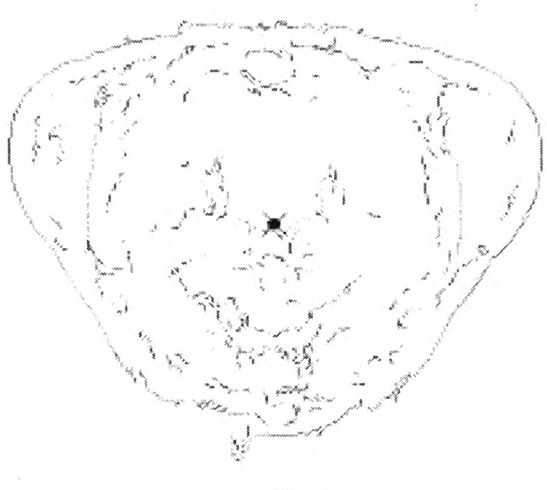


e111

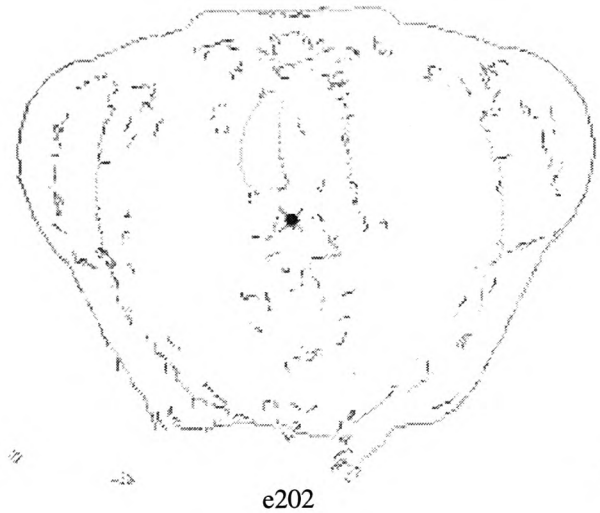
Species 2 : *Tetramesa hyalipennis*



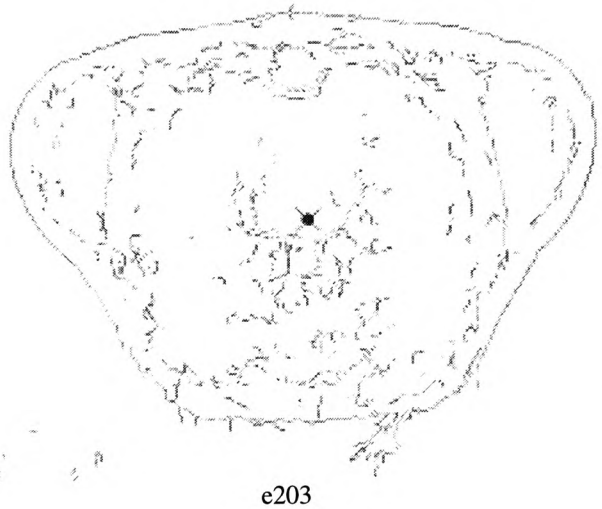
e200



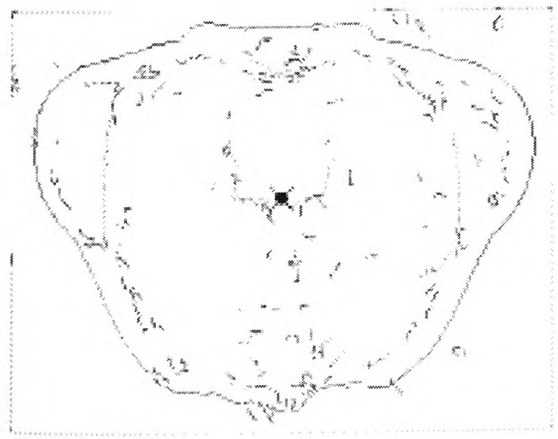
e201



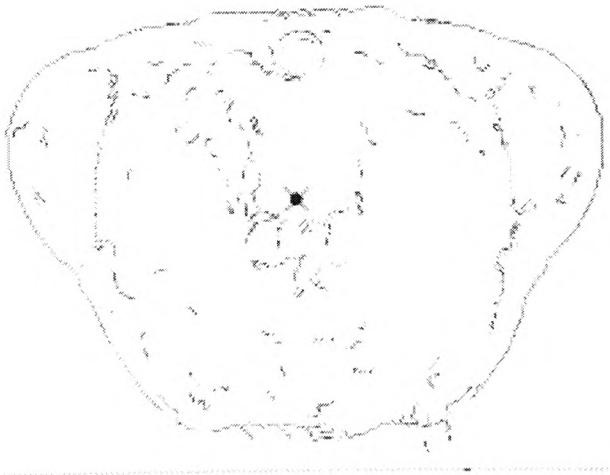
e202



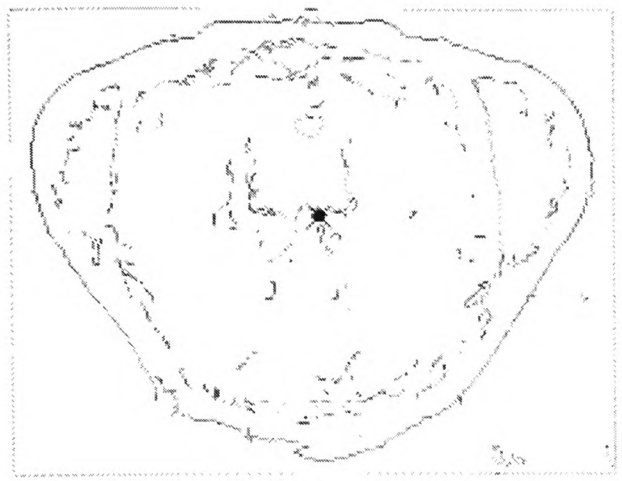
e203



e204

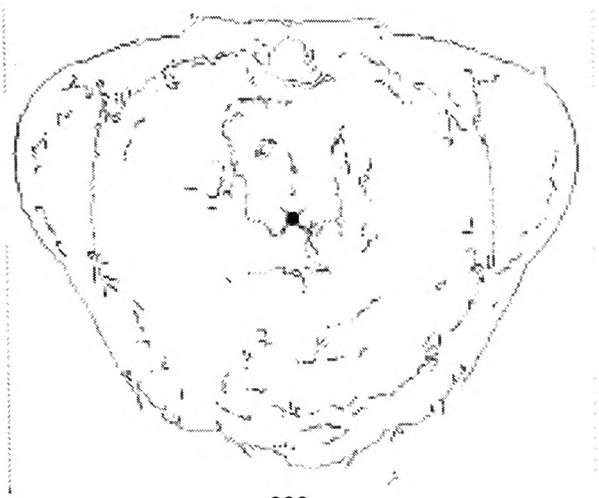


e205

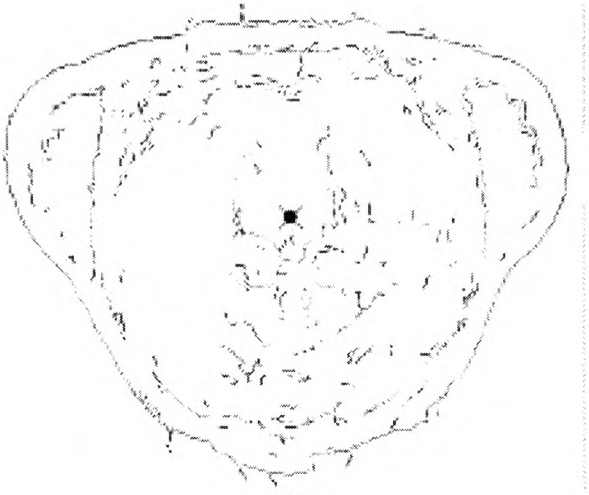


e206

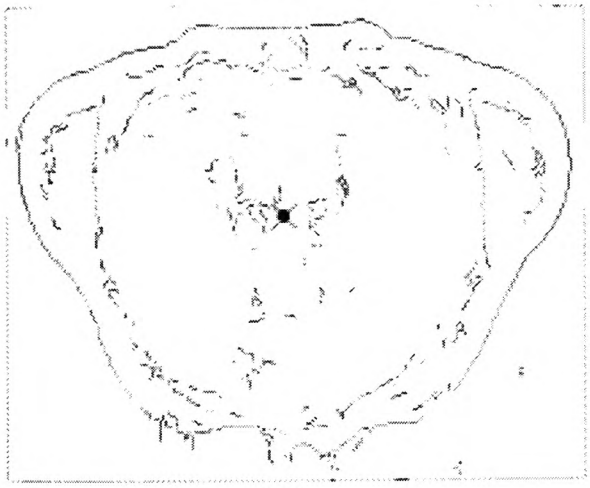
Species 3 : *Tetramesa calamagrostidis*



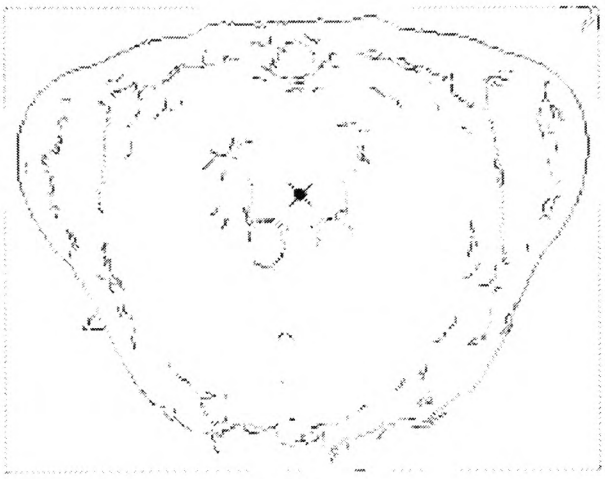
e300



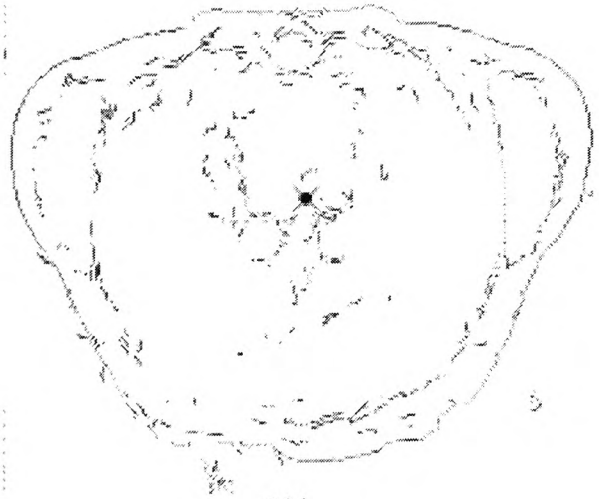
e301



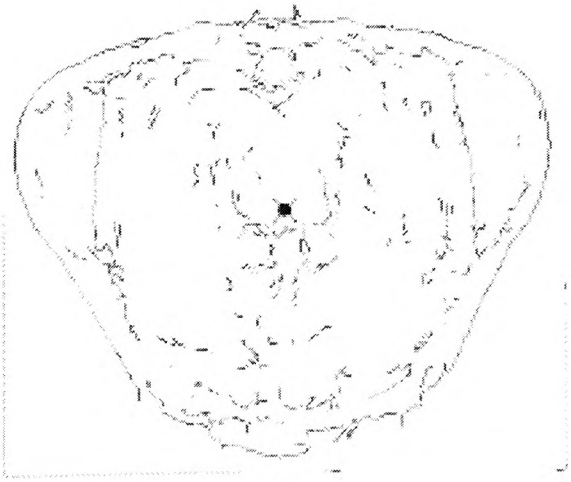
e302



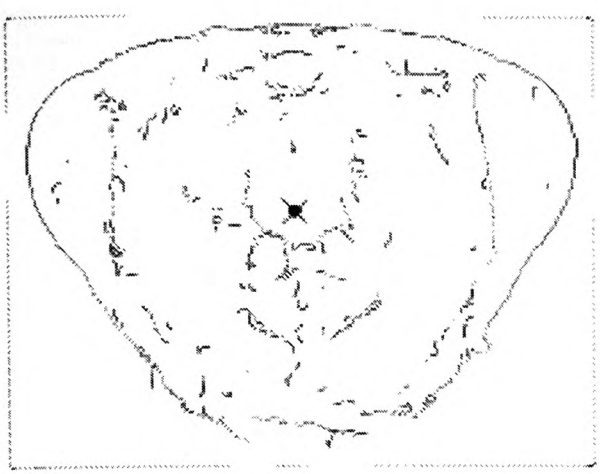
e303



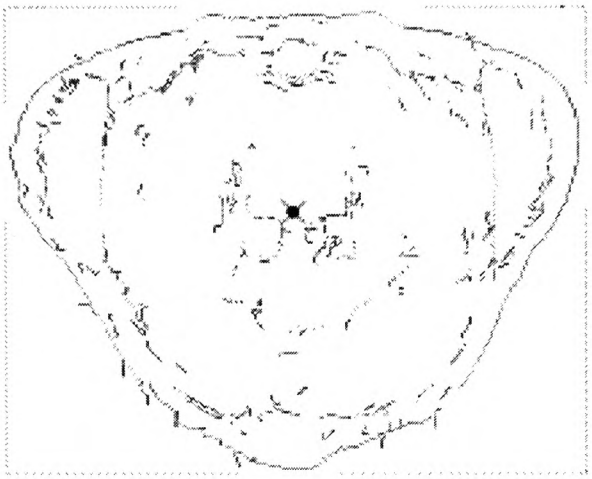
e304



e305

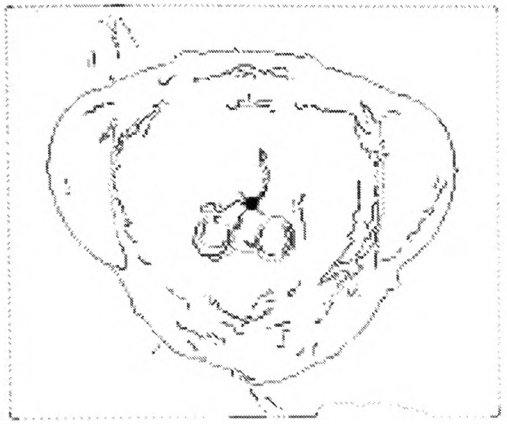


e306

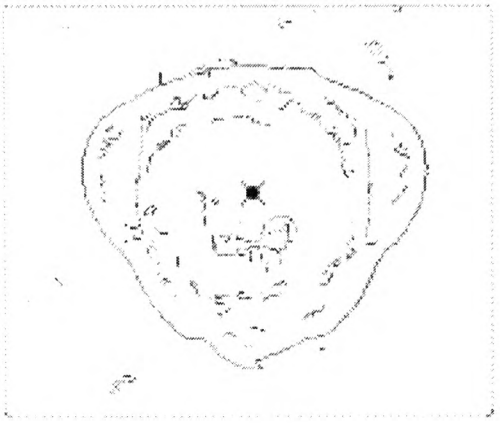


e307

Species 4 : *Tetramesa fulvicollis*

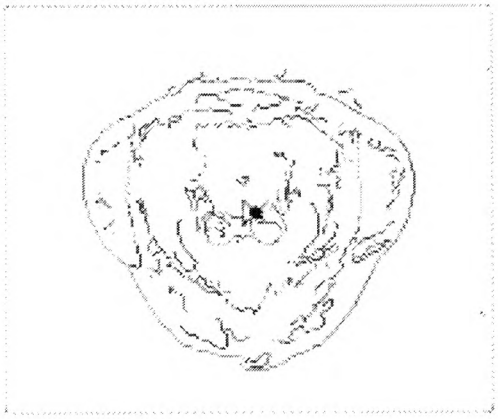


e400

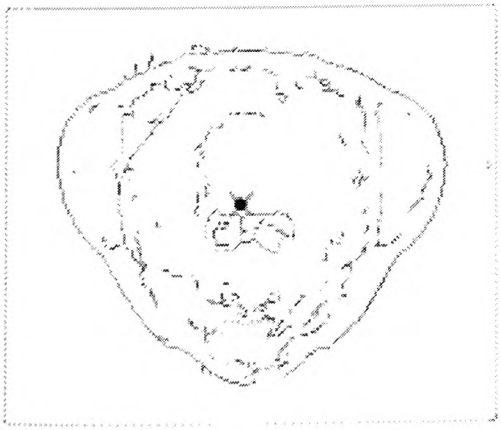


e401

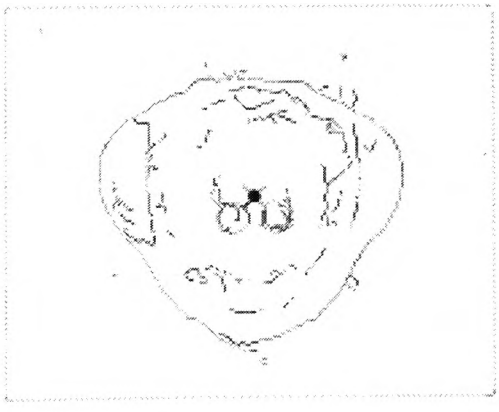
e402



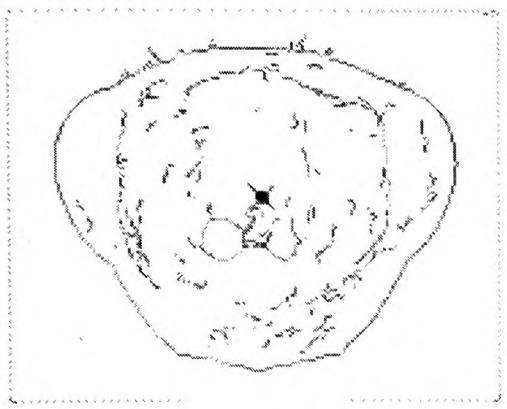
e403



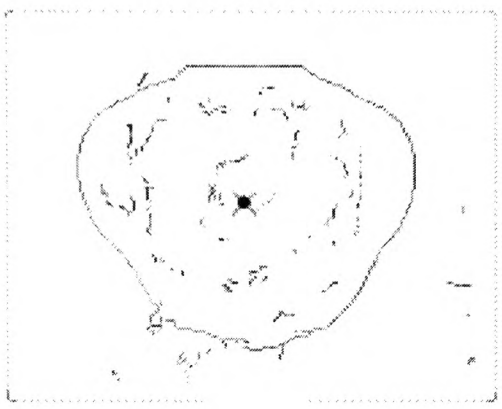
e404



e405



e406

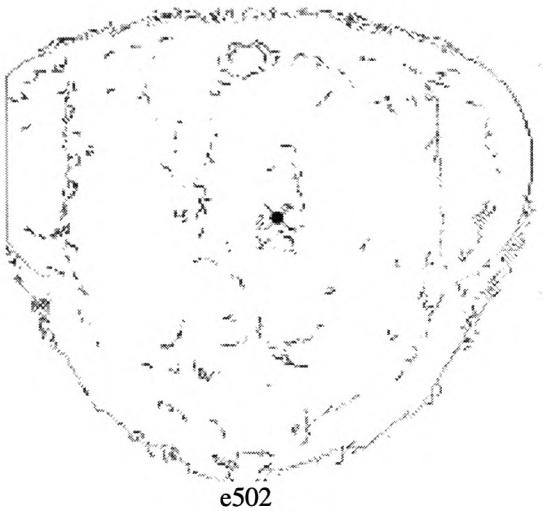


e407

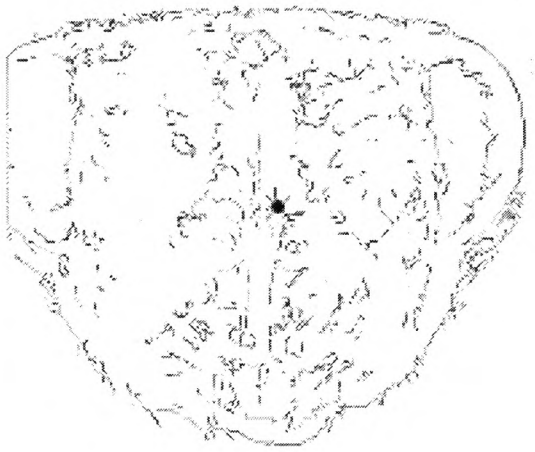
Species 5 : *Tetramesa eximia*

e500

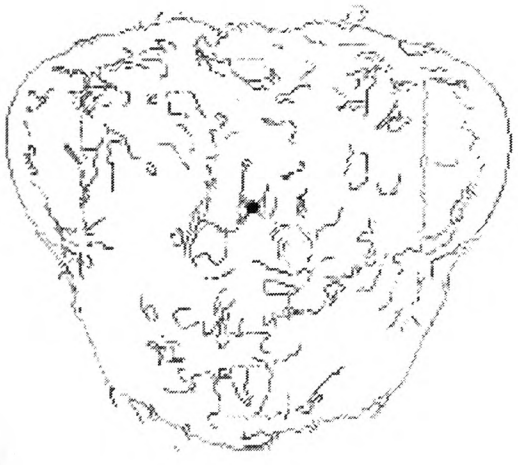
e501



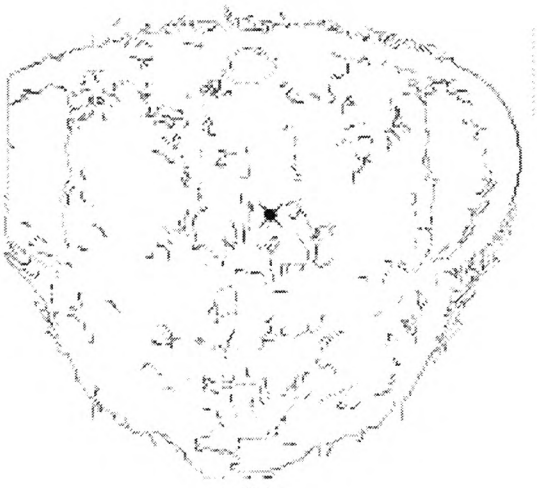
e502



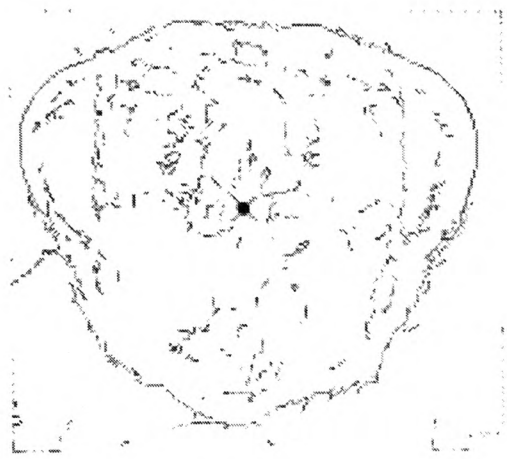
e503



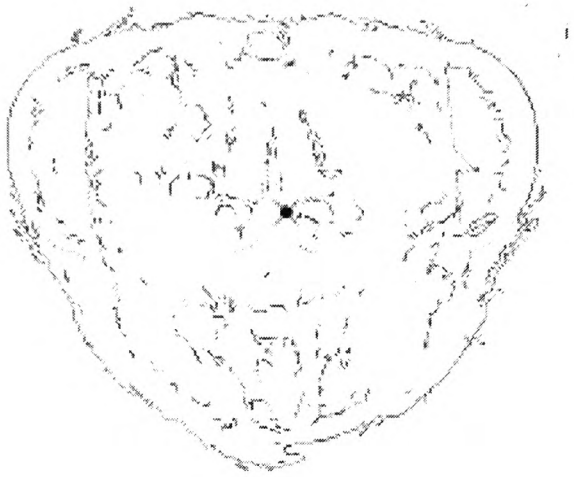
e504



e505



e506

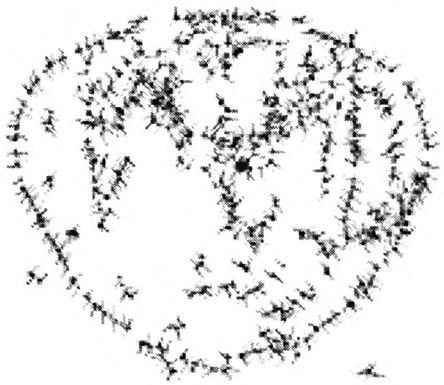


e507

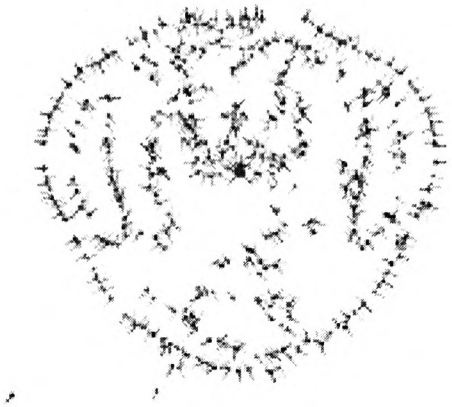
Appendix 6

Dominant Curvature Point Maps

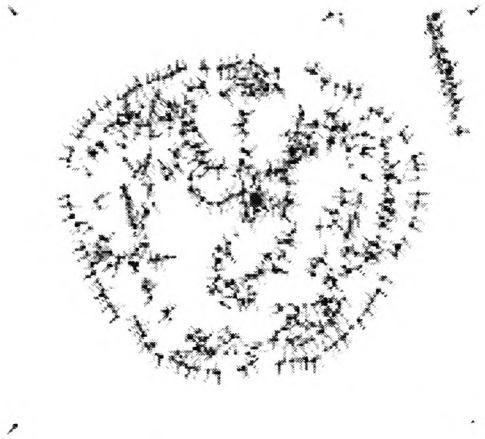
Species 1 : *Tetramesa linearis*



e100

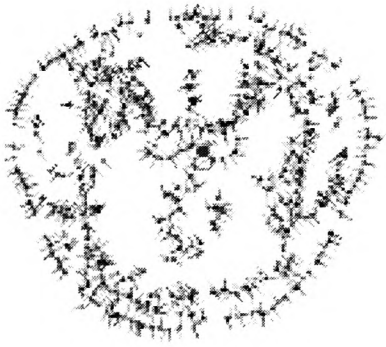


e101

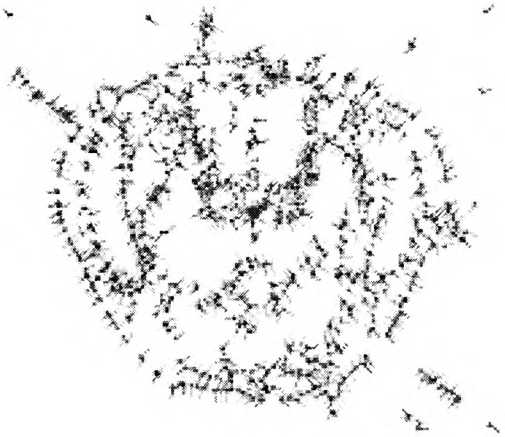


e102

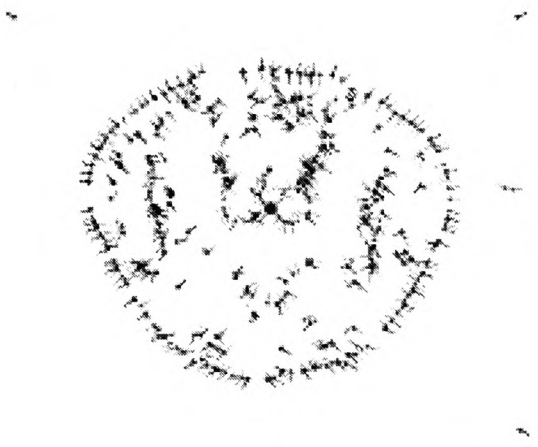
e103



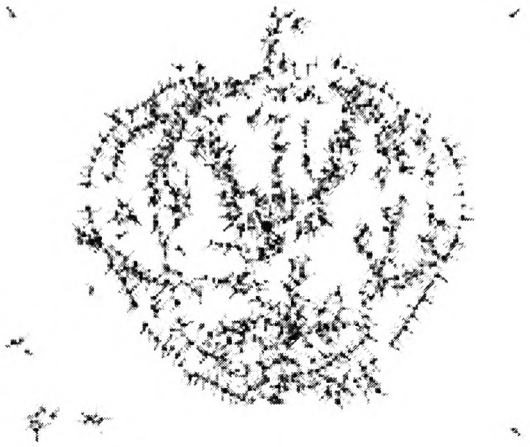
e104



e105



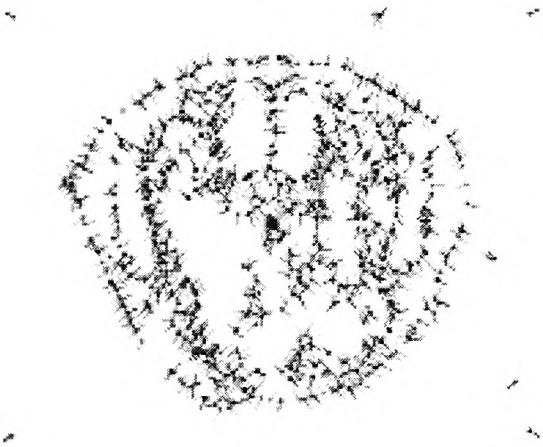
e106



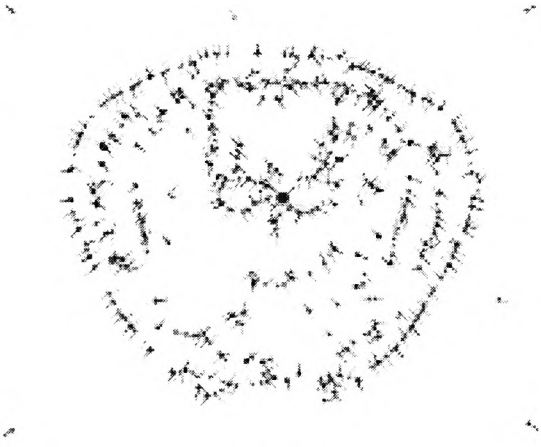
e107

e108

e109



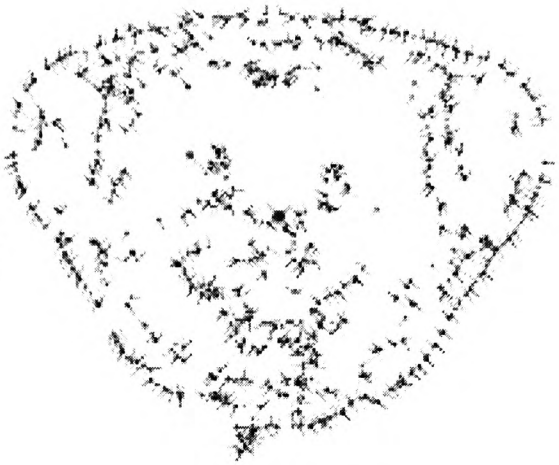
e110



e111

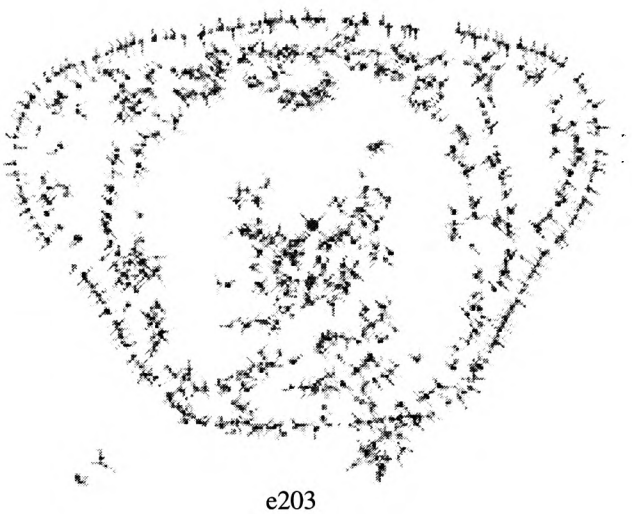
Species 2 : *Tetramesa hyalipennis*

e200

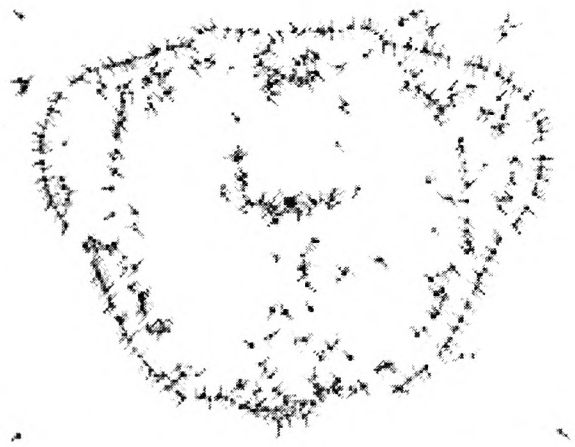


e201

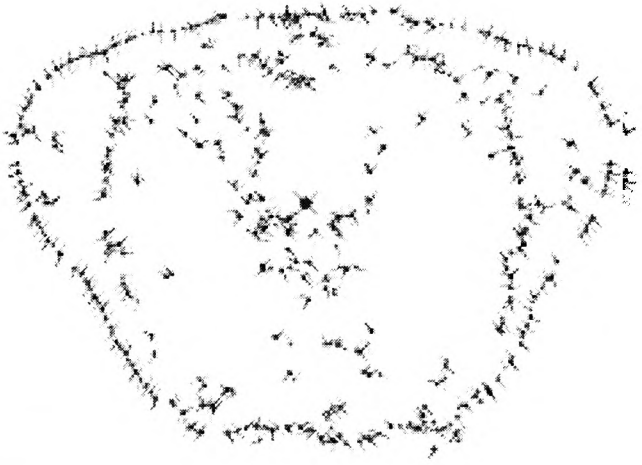
e202



e203



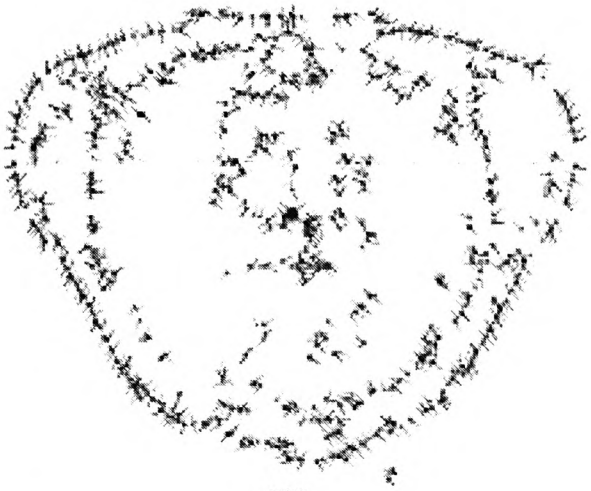
e204



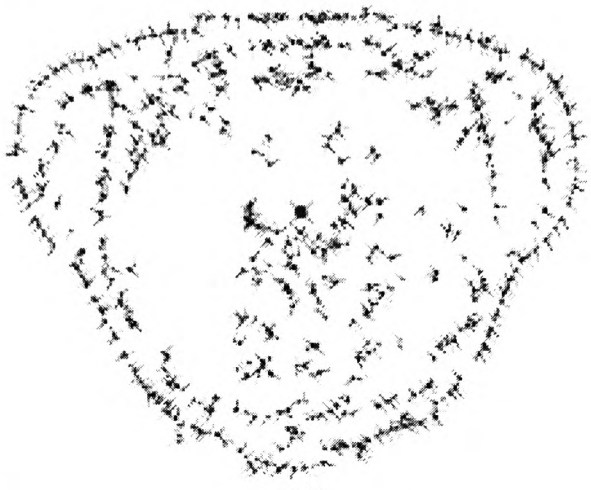
e205

e206

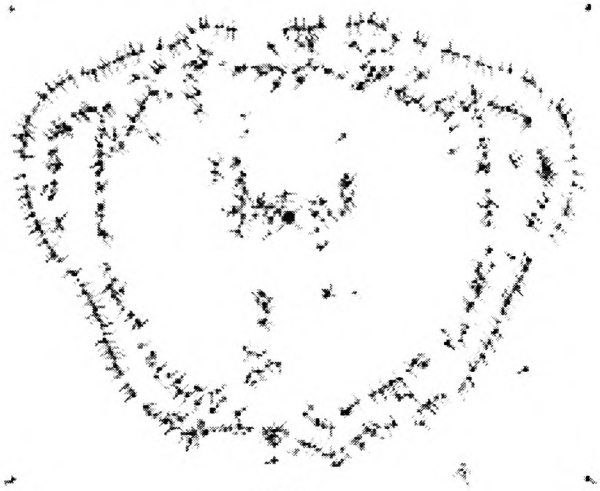
Species 3 : *Tetramesa calamagrostidis*



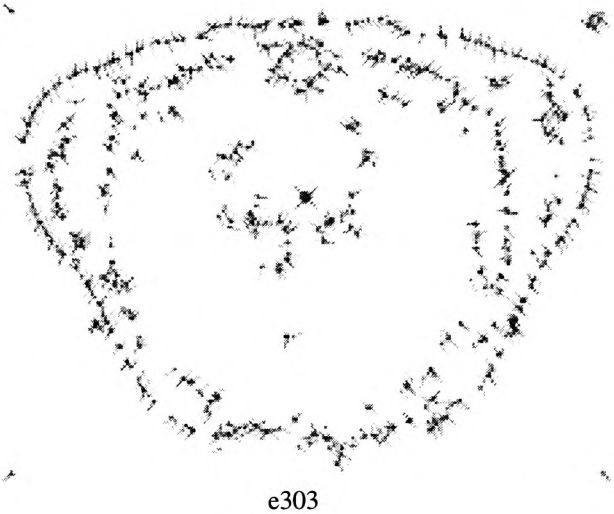
e300



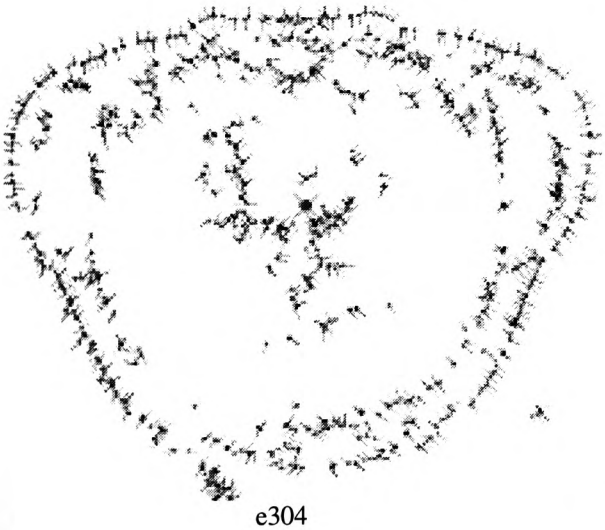
e301



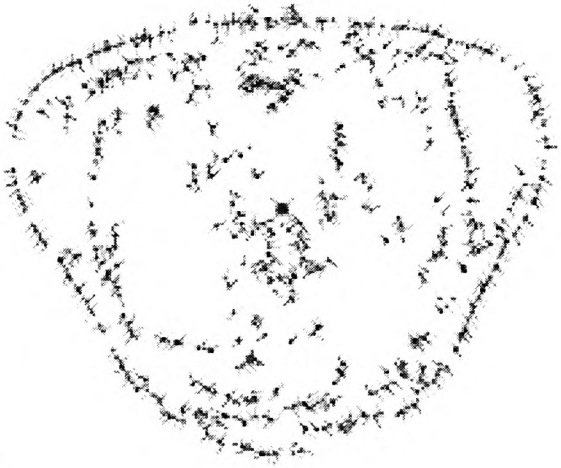
e302



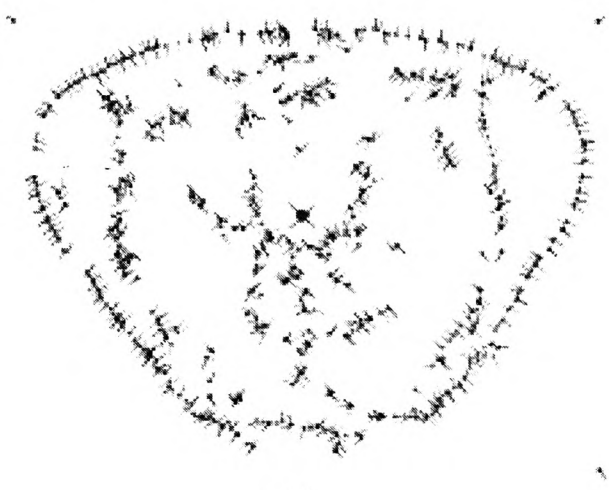
e303



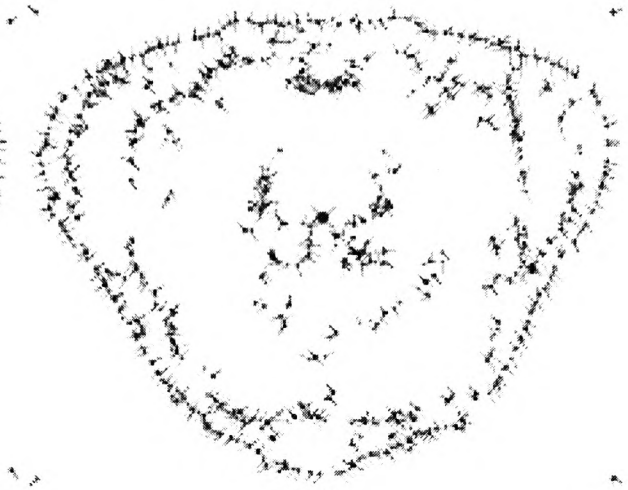
e304



e305

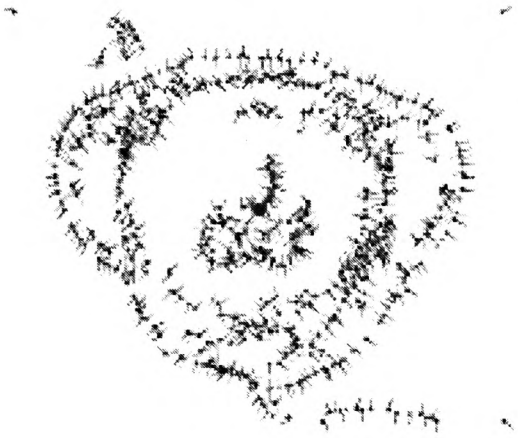


e306



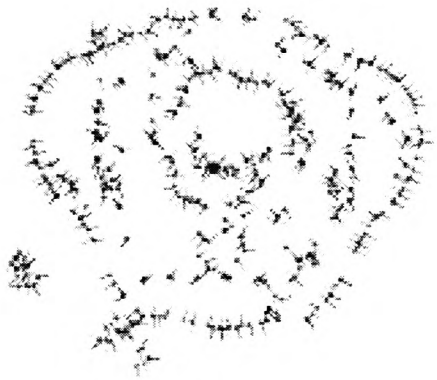
e307

Species 4 : *Tetramesa fulvicollis*

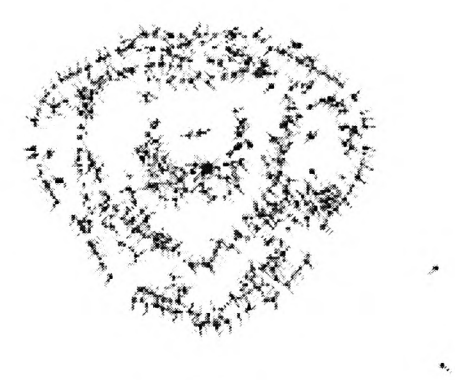


e400

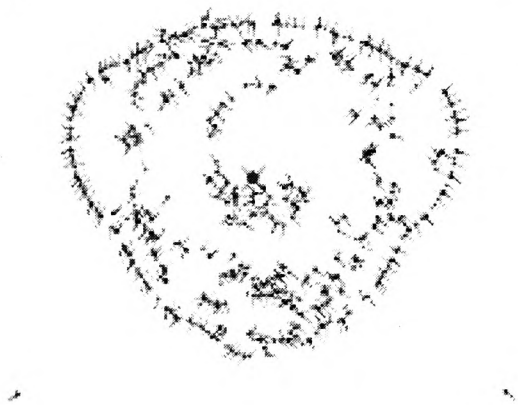
e401



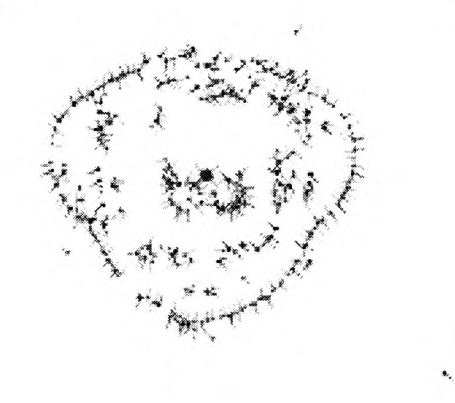
e402



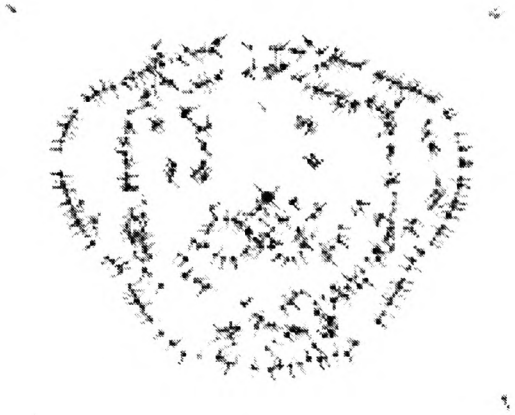
e403



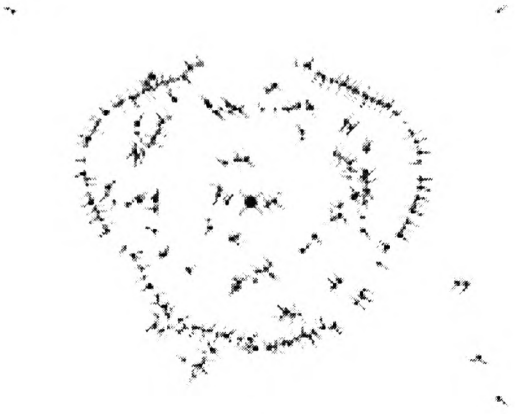
e404



e405

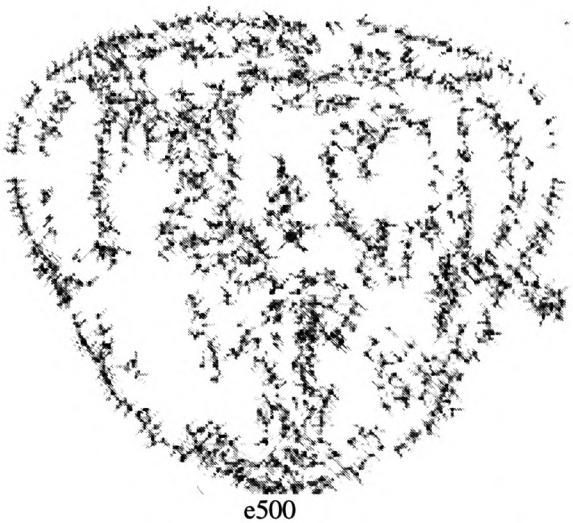


e406

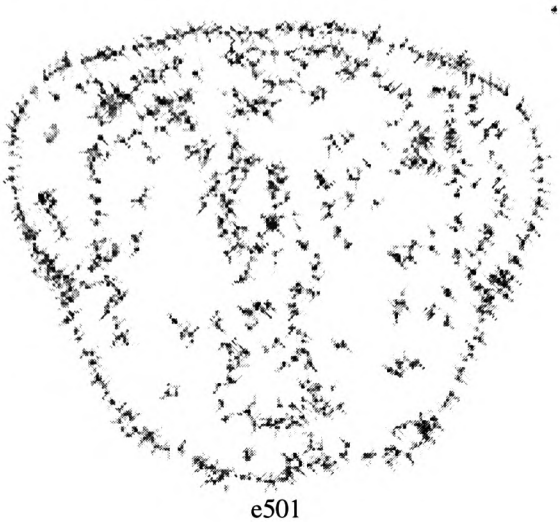


e407

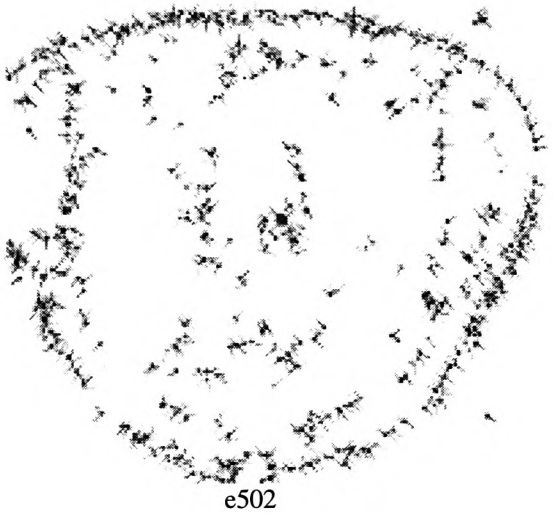
Species 5 : *Tetramesa eximia*



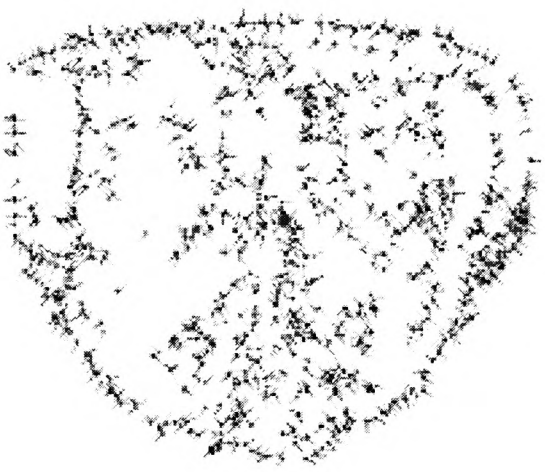
e500



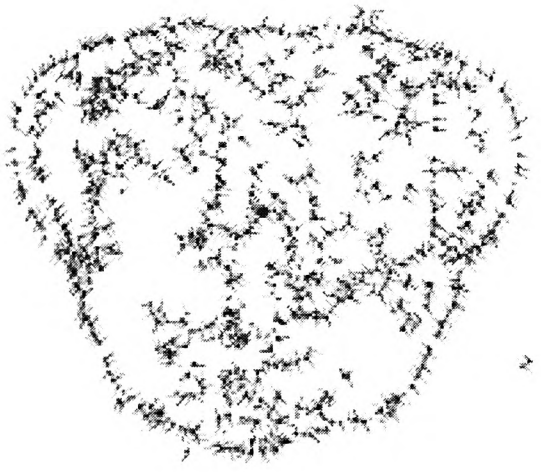
e501



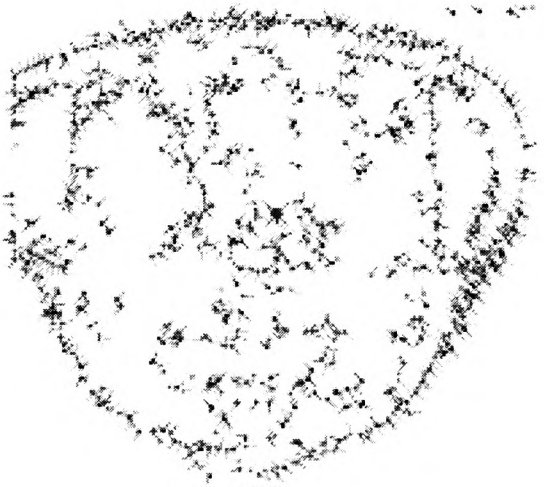
e502



e503

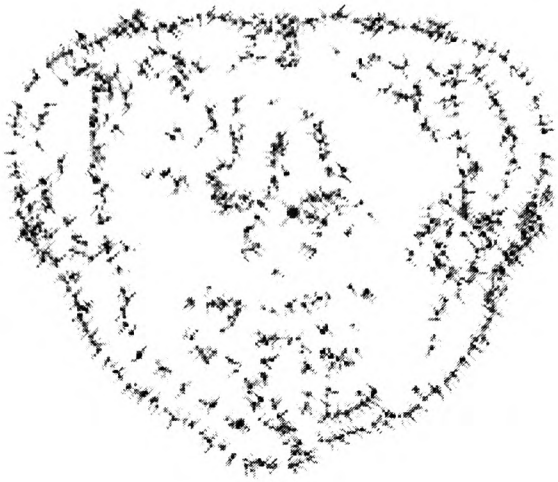


e504



e505

e506

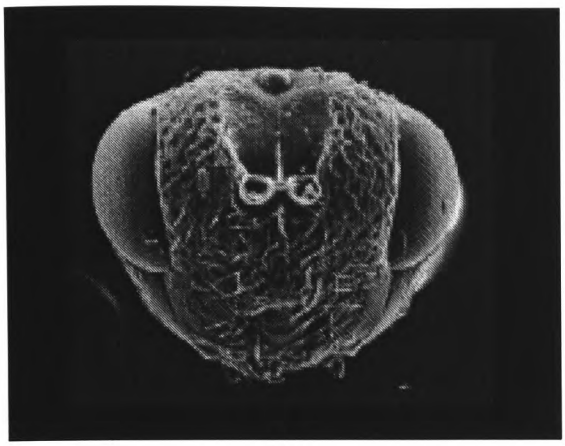


e507

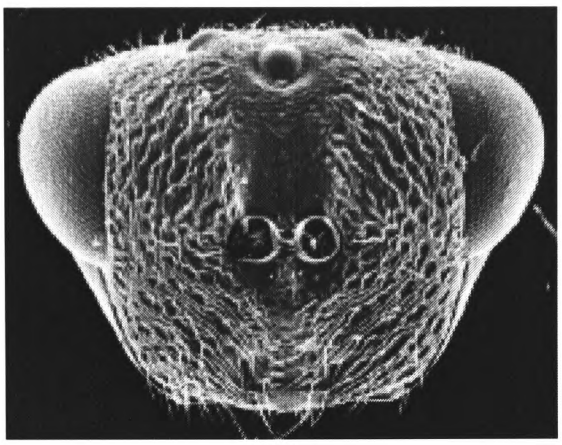
Appendix 7

Wasp Head Model Training Data

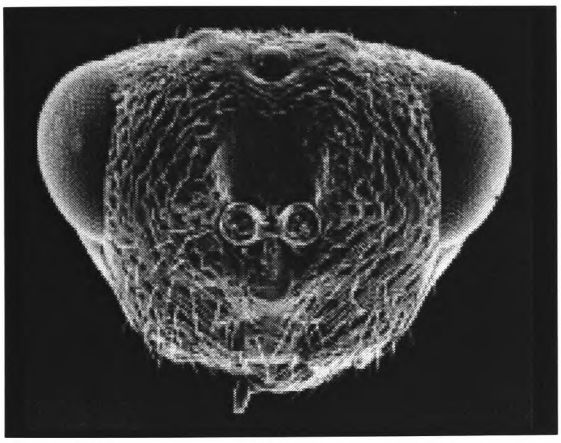
Training Instances



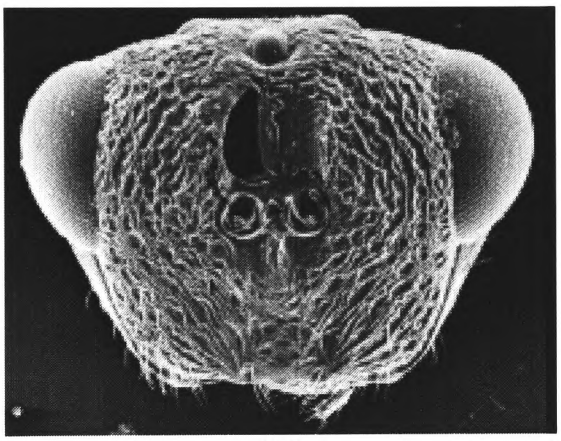
e100



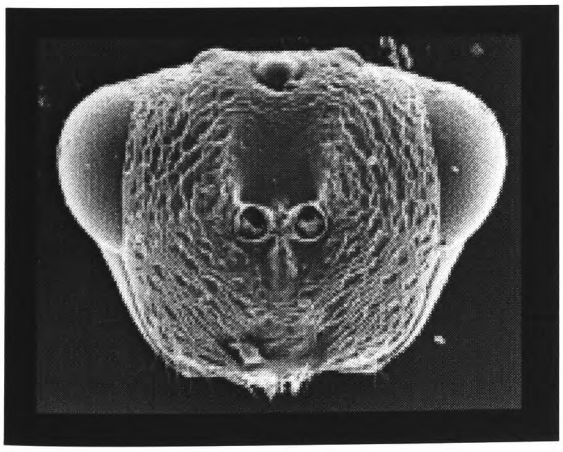
e200



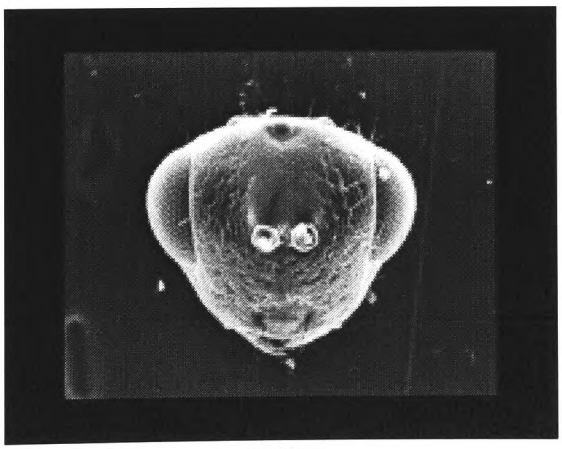
e201



e202



e204

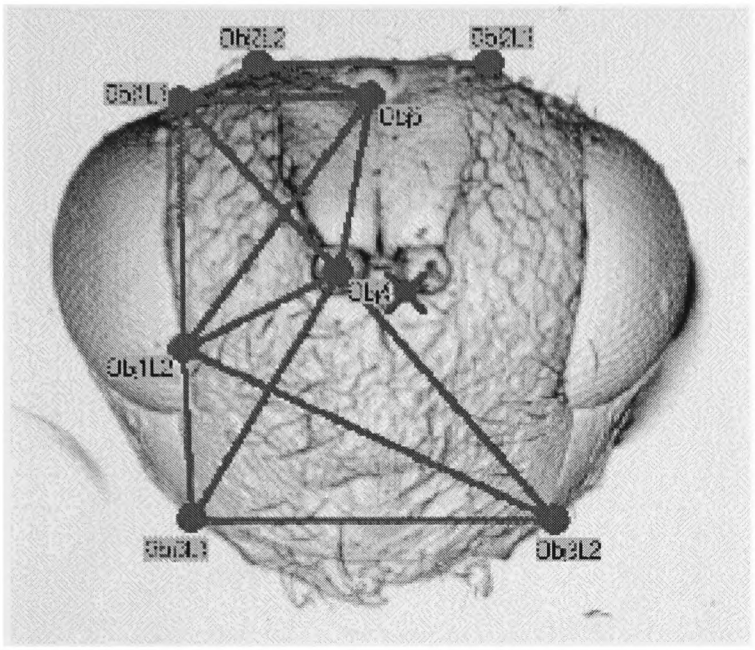


e405

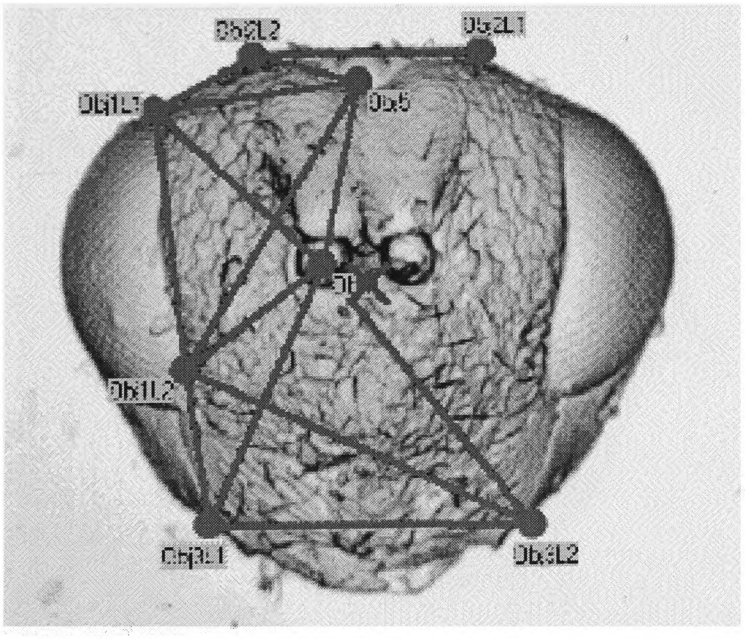
Appendix 8

Landmark Results

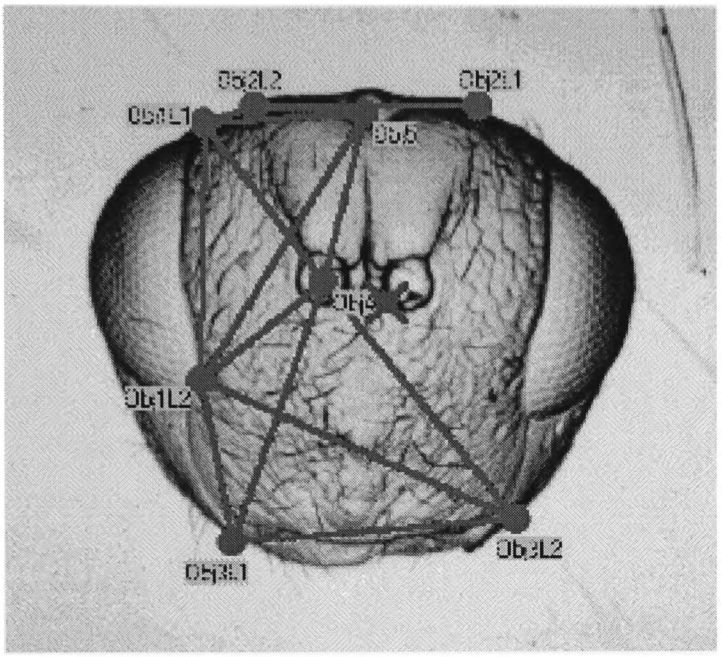
Species 1 : *Tetramesa linearis*



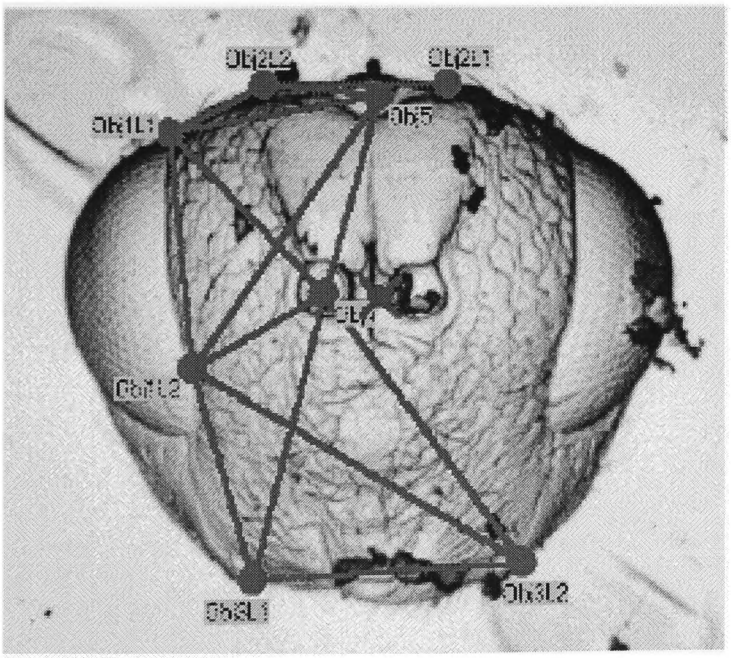
e100



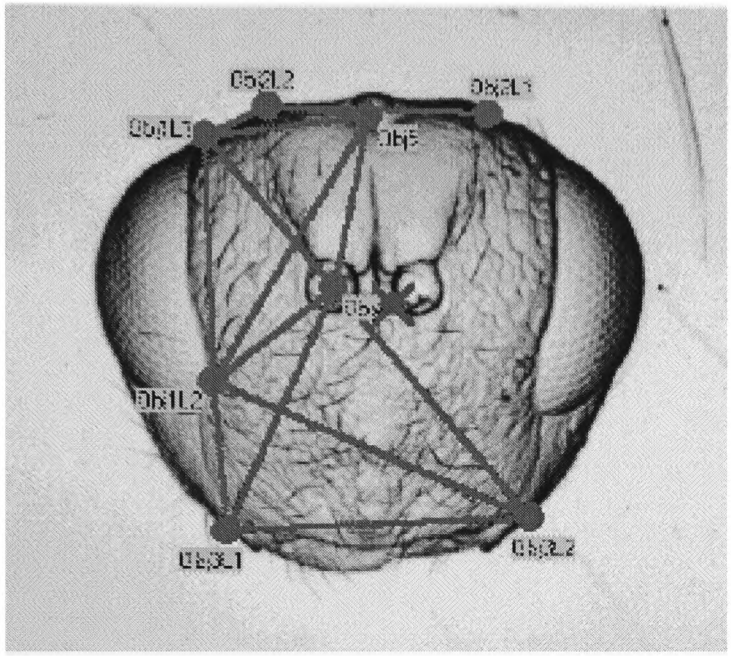
e101


Ob2L2
Ob2L1
Ob1L1
Ob5
Ob4
Ob1L2
Ob3L2
Ob3L1

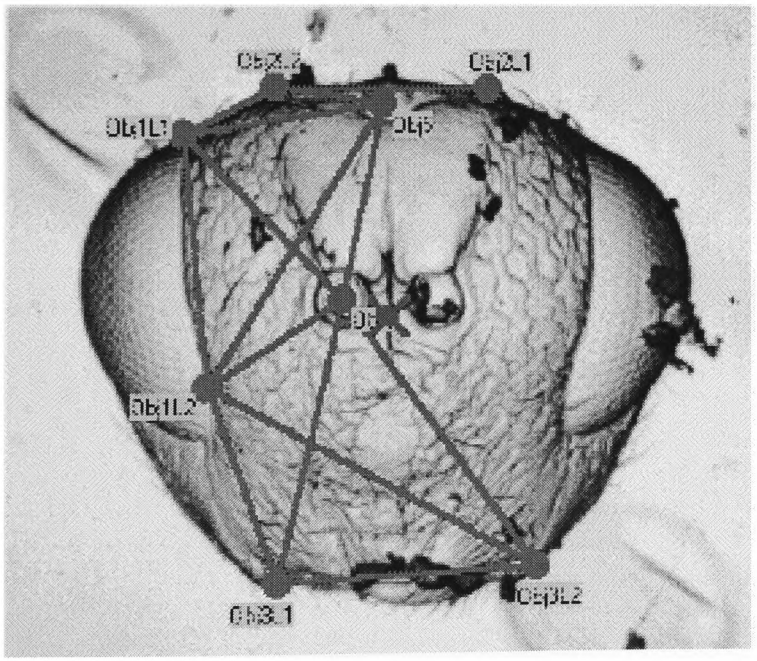
e102


Ob2L2
Ob2L1
Ob1L1
Ob5
Ob4
Ob1L2
Ob3L2
Ob3L1

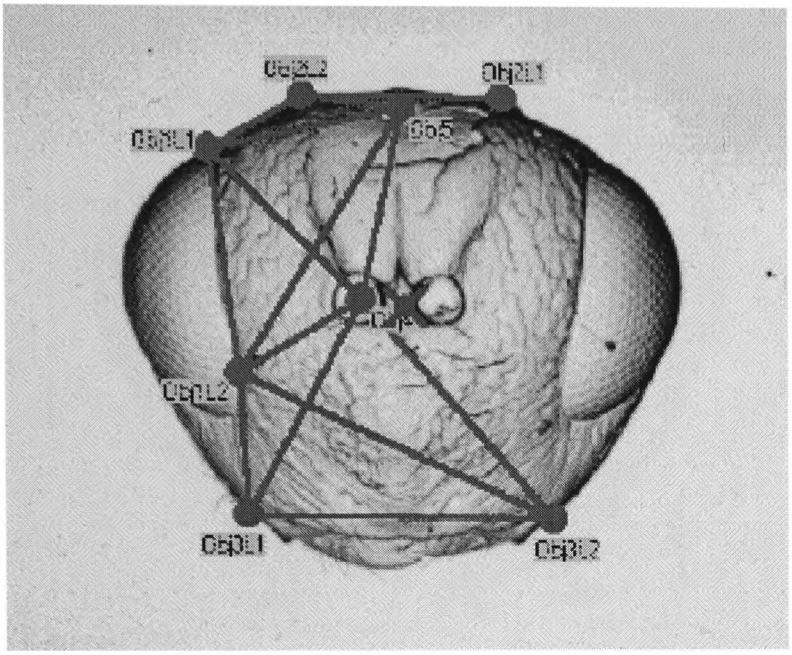
e103



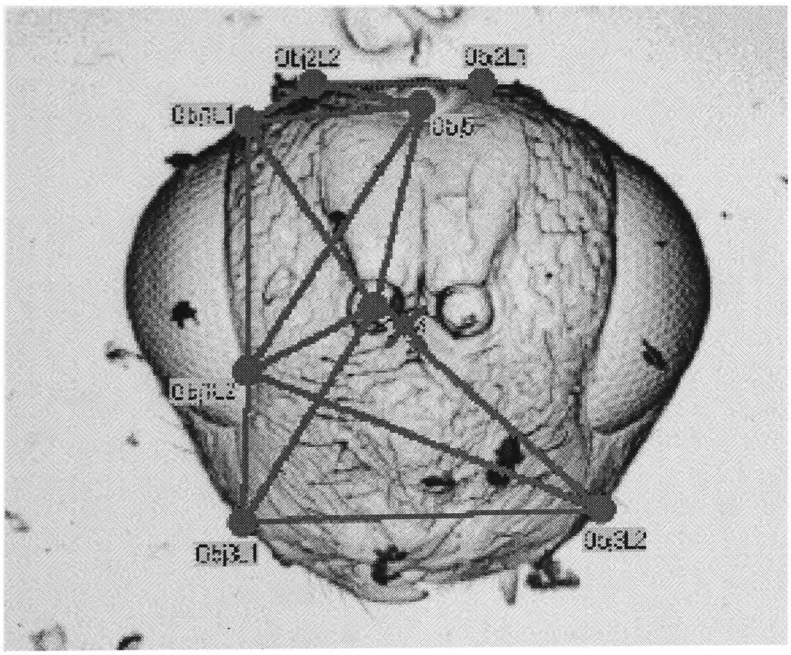
e104



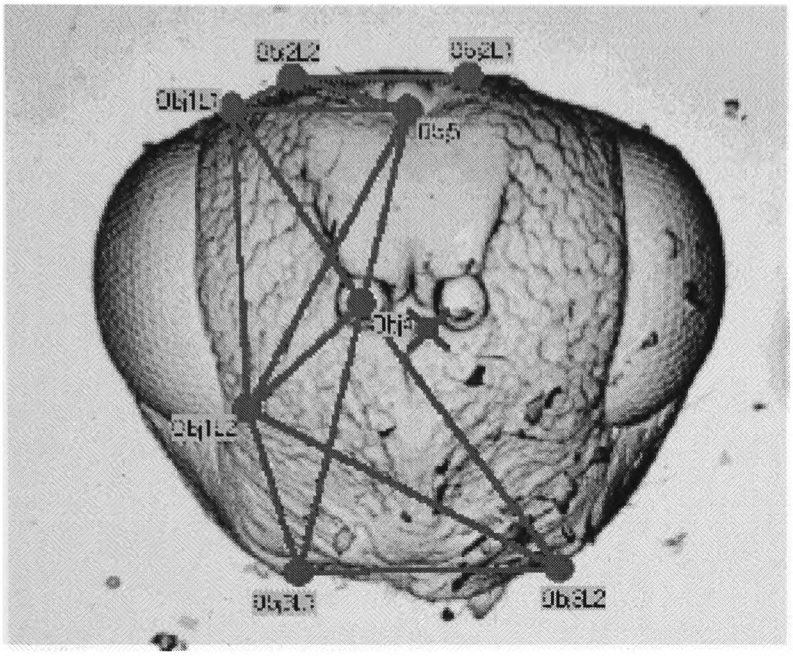
e105


Obj2L2
Obj2L1
Obj5
Obj1L1
Obj1L2
Obj3L1
Obj3L2

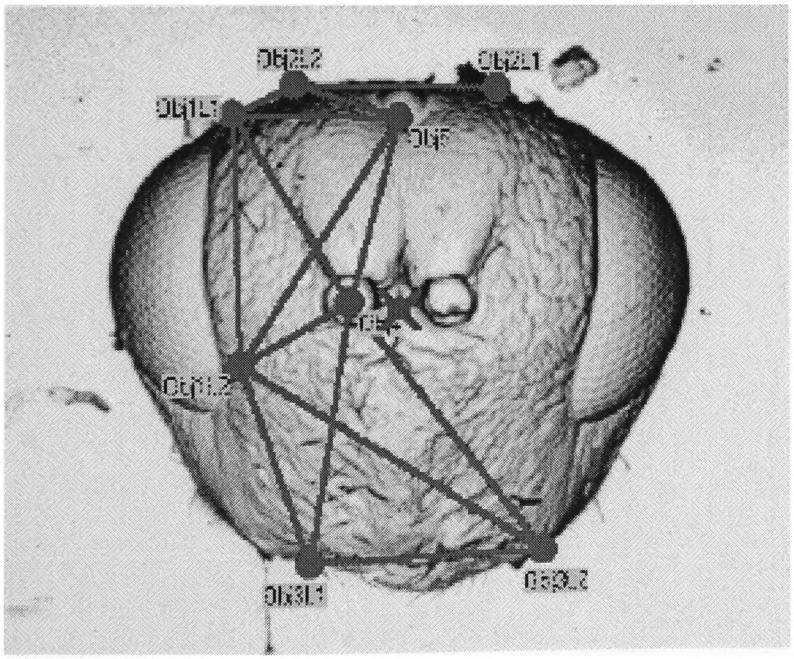
e106


Obj2L2
Obj2L1
Obj1L1
Obj5
Obj1L2
Obj3L1
Obj3L2

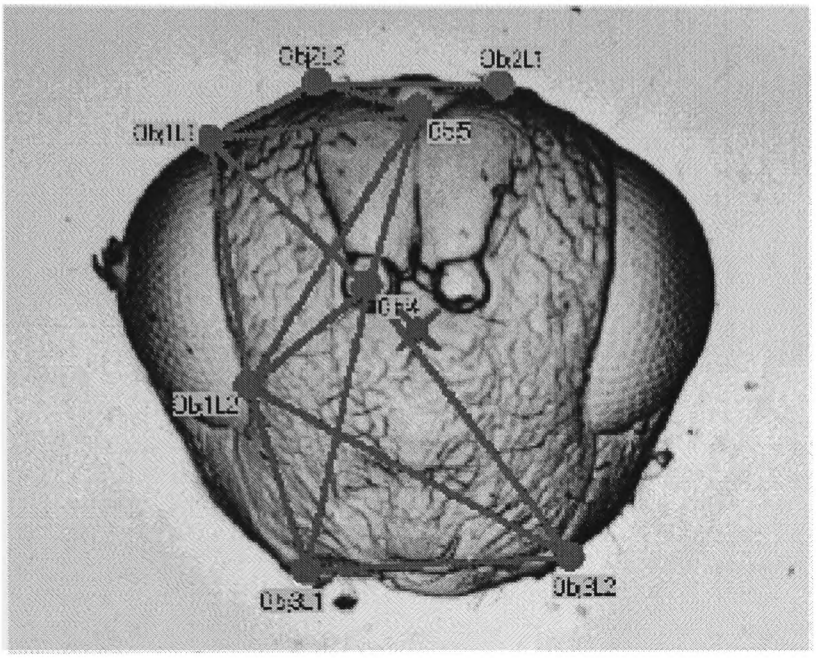
e107


Ob2L2
Ob2L1
Ob1L1
Ob5
Ob4
Ob1L2
Ob3L1
Ob3L2

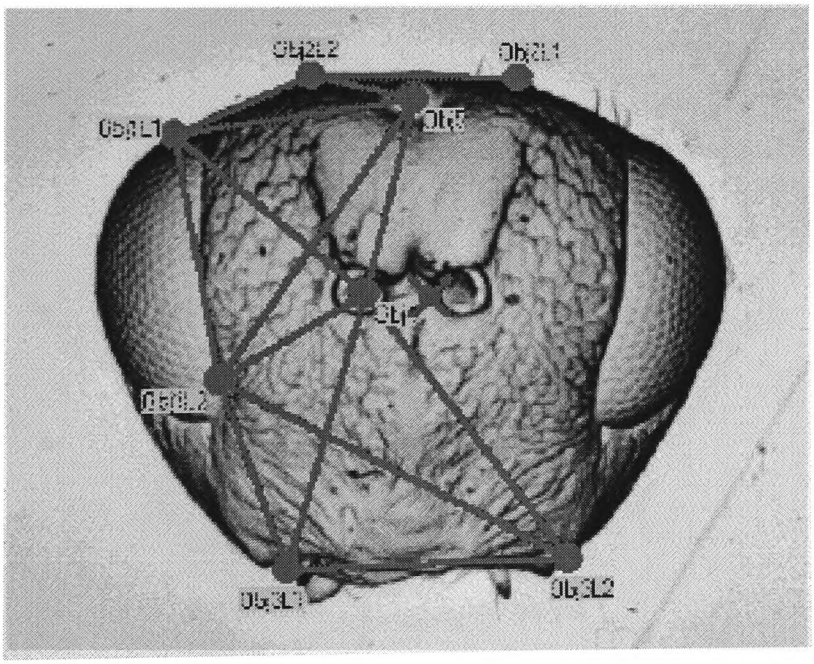
e108


Ob2L2
Ob2L1
Ob1L1
Ob5
Ob4
Ob1L2
Ob3L1
Ob3L2

e109

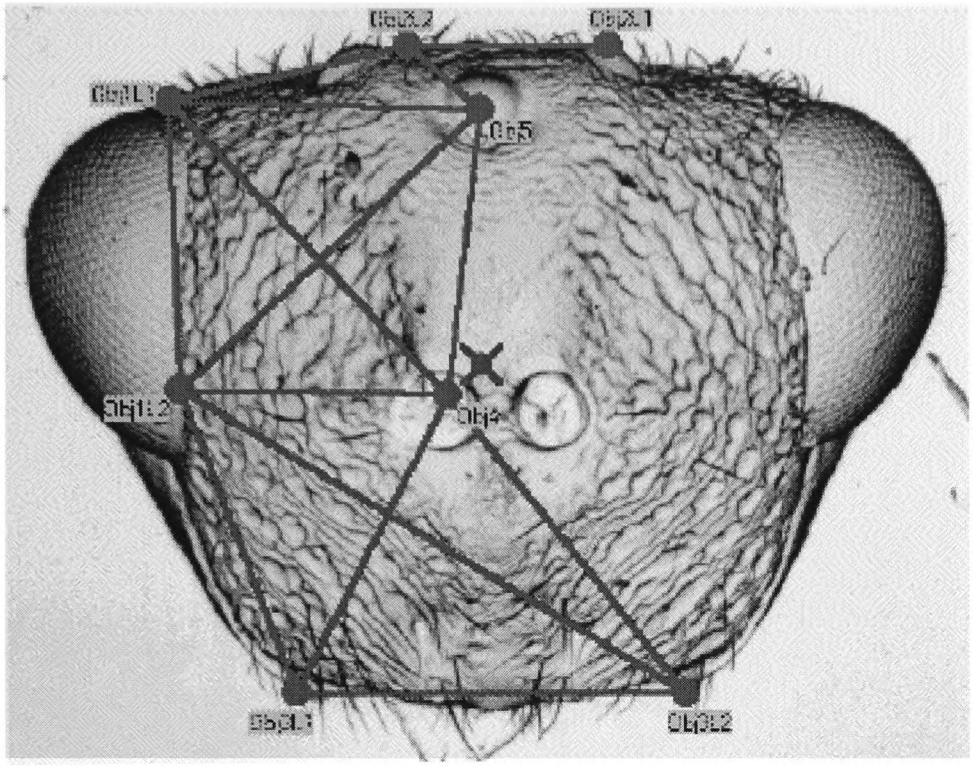

Ob2L2
Ob2L1
Ob1L1
Ob5
Ob4
Ob1L2
Ob3L1
Ob3L2

e110

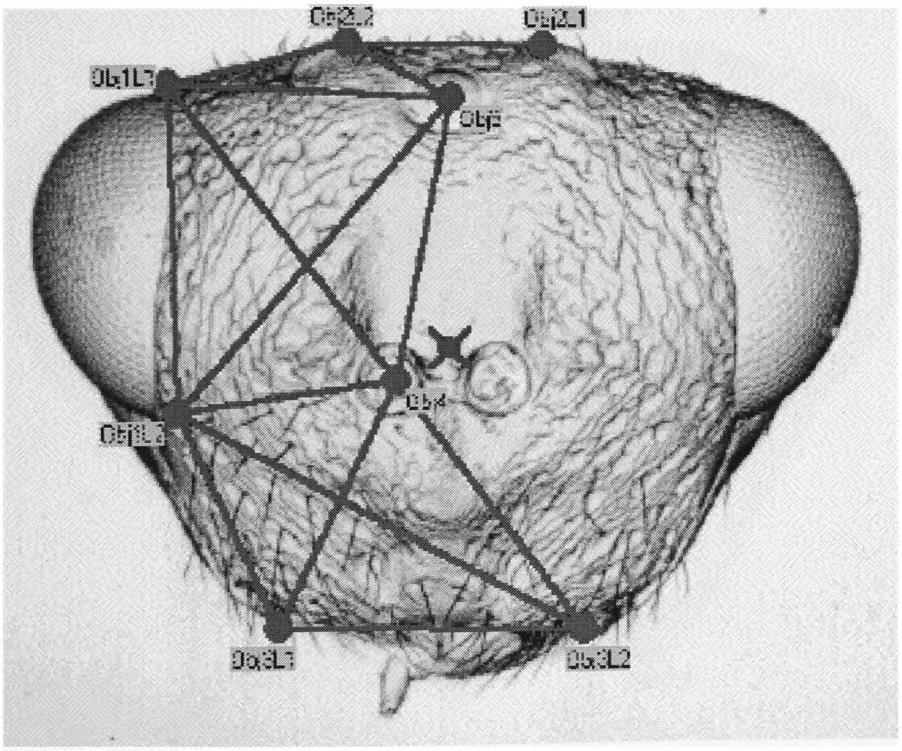

Ob2L2
Ob2L1
Ob1L1
Ob5
Ob4
Ob3L2
Ob3L1
Ob3L2

e111

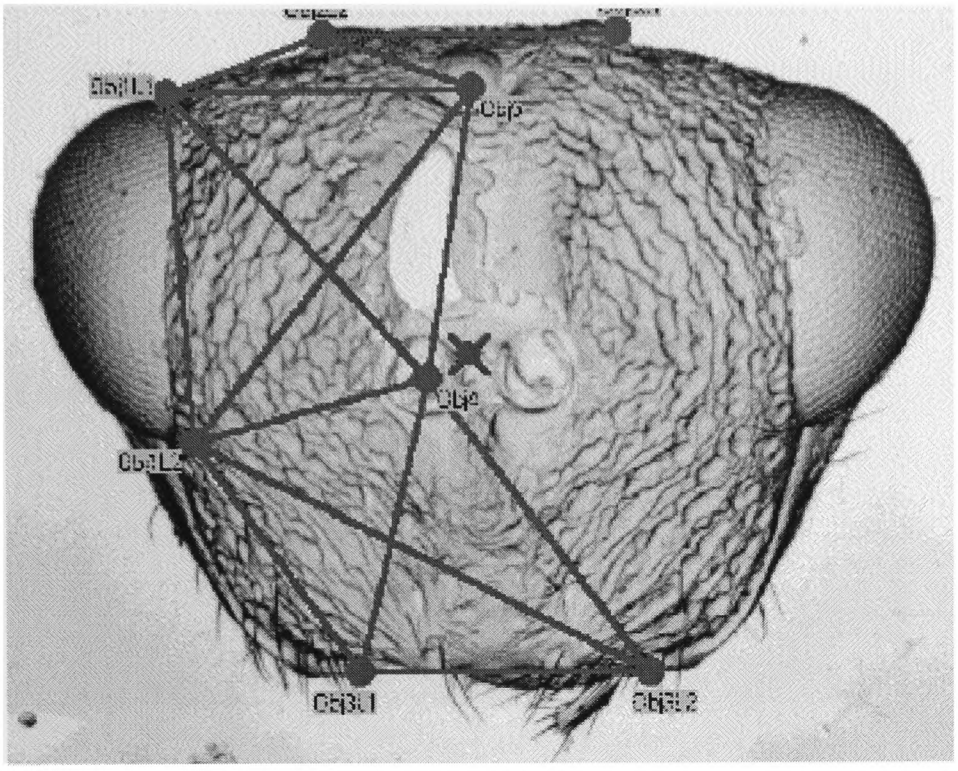
Species 2 : *Tetramesa hyalipennis*



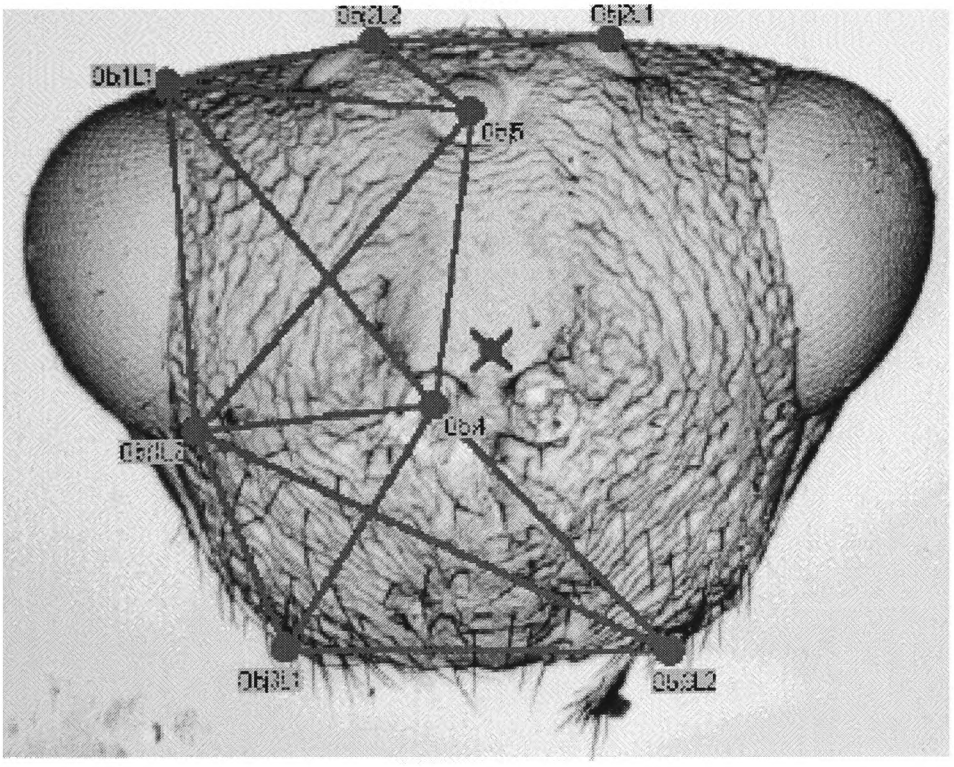
e200



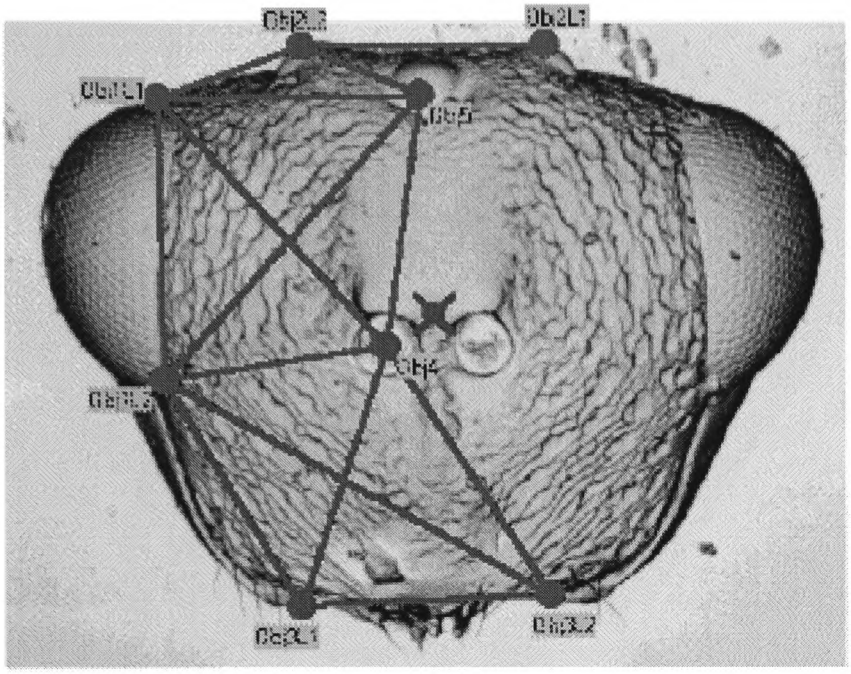
e201


Obj5
Obj4
Obj3L1
Obj3L2

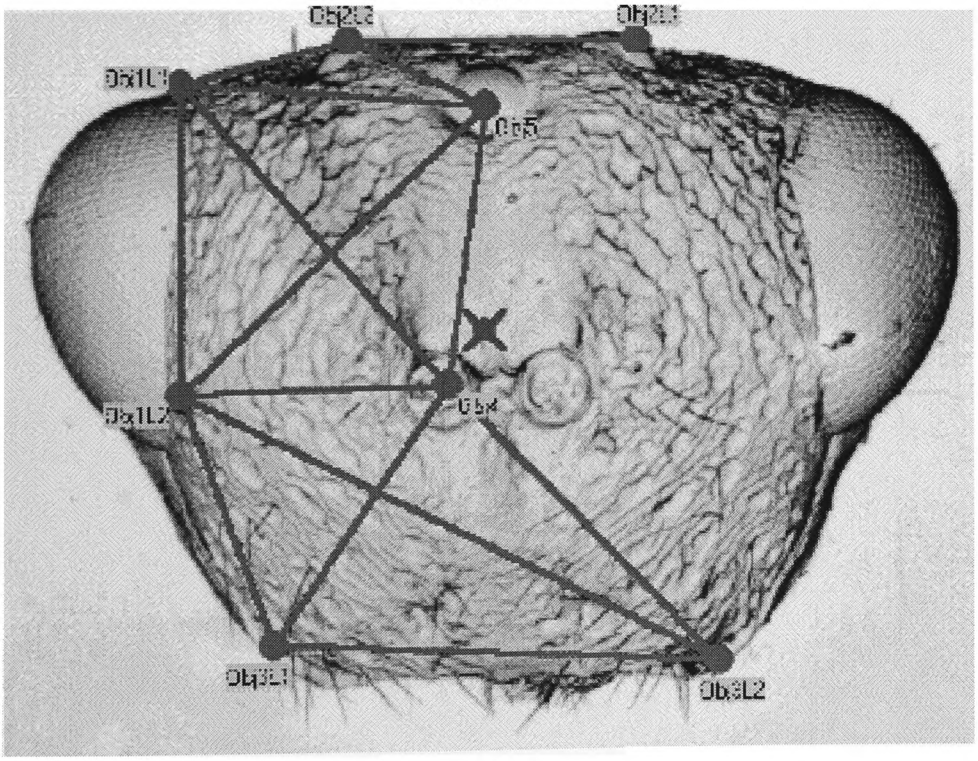
e202


Obj2L2
Obj2L1
Ob1L1
Obj5
Obj4
Obj3L1
Obj3L2

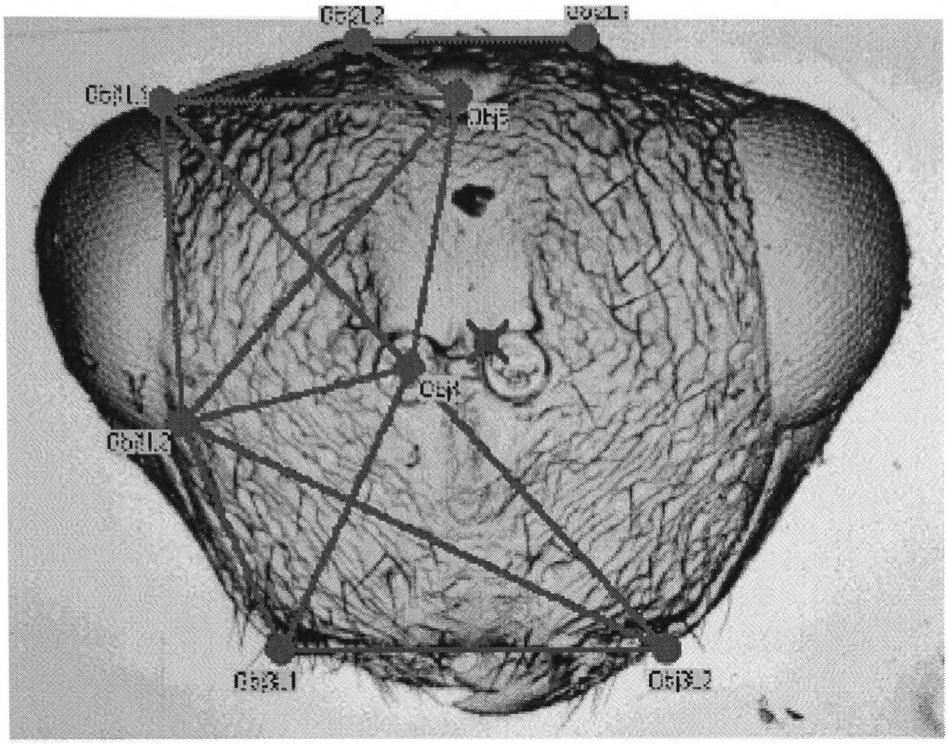
e203


Obj2L2
Obj2L1
Obj5
Obj4
Obj3L1
Obj3L2

e204

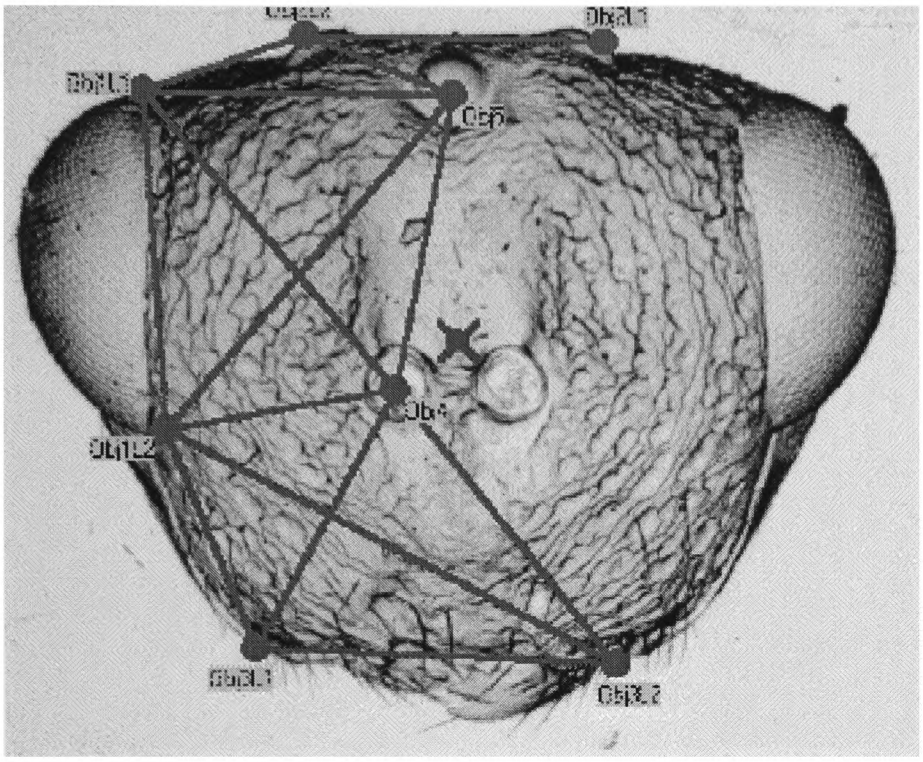

Obj2L2
Obj2L1
Obj5
Obj4
Obj3L1
Obj3L2

e205

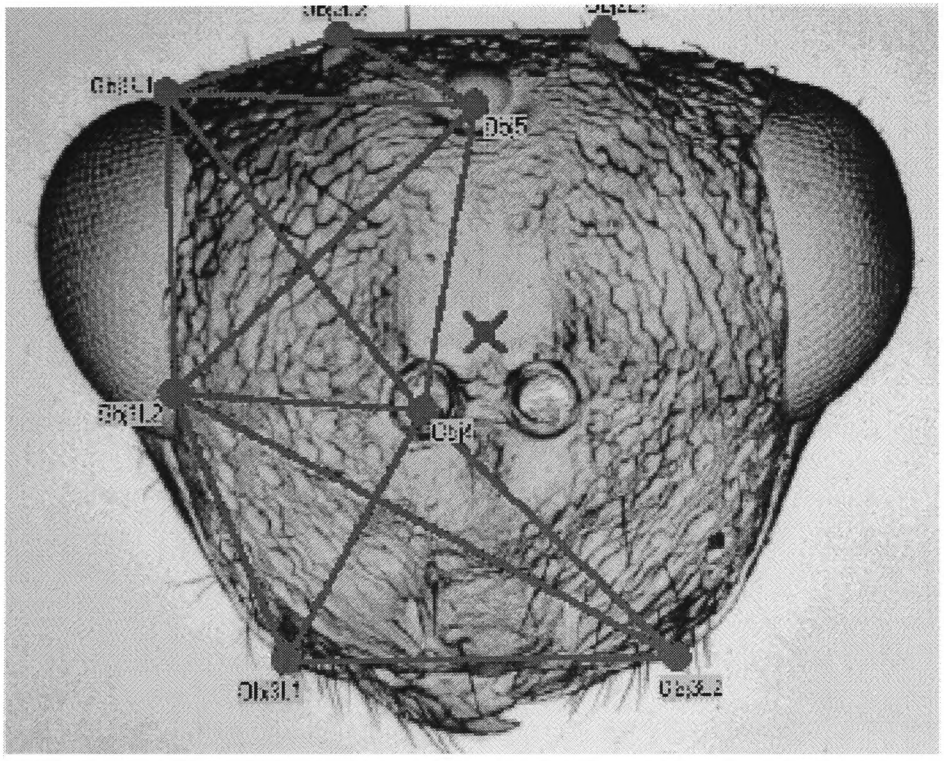


e206

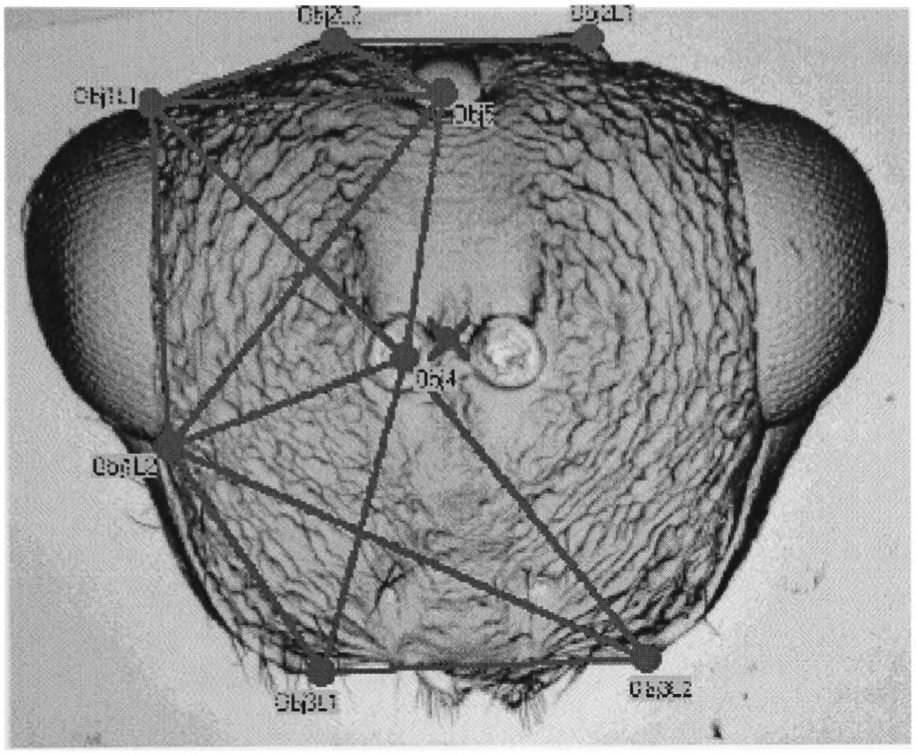
Species 3 : *Tetramesa calamagrostidis*



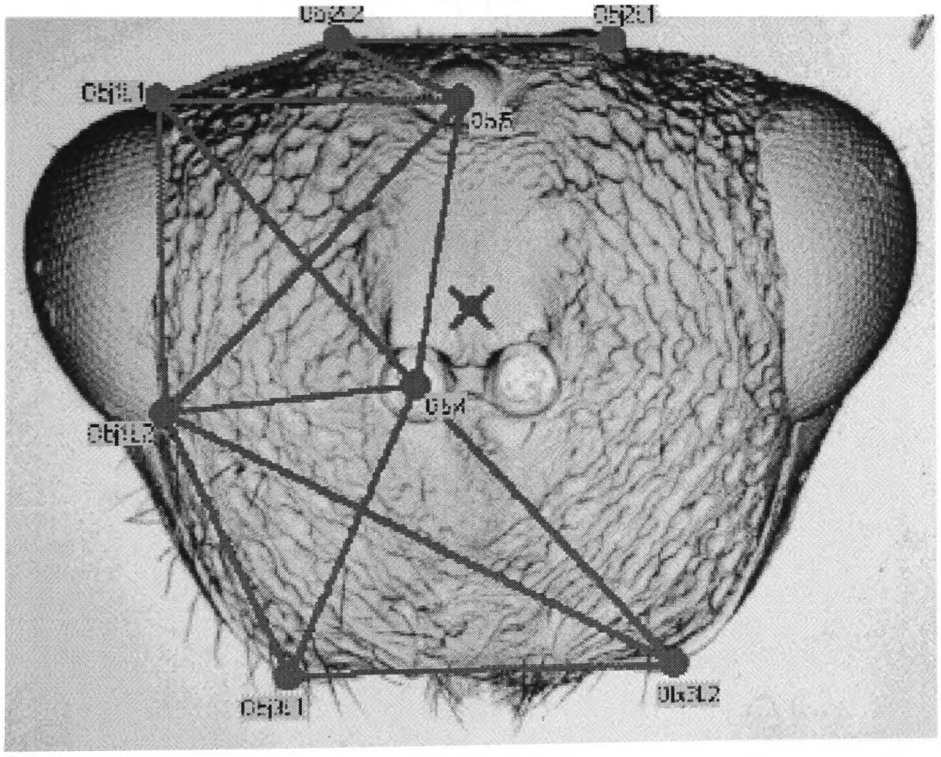
e300



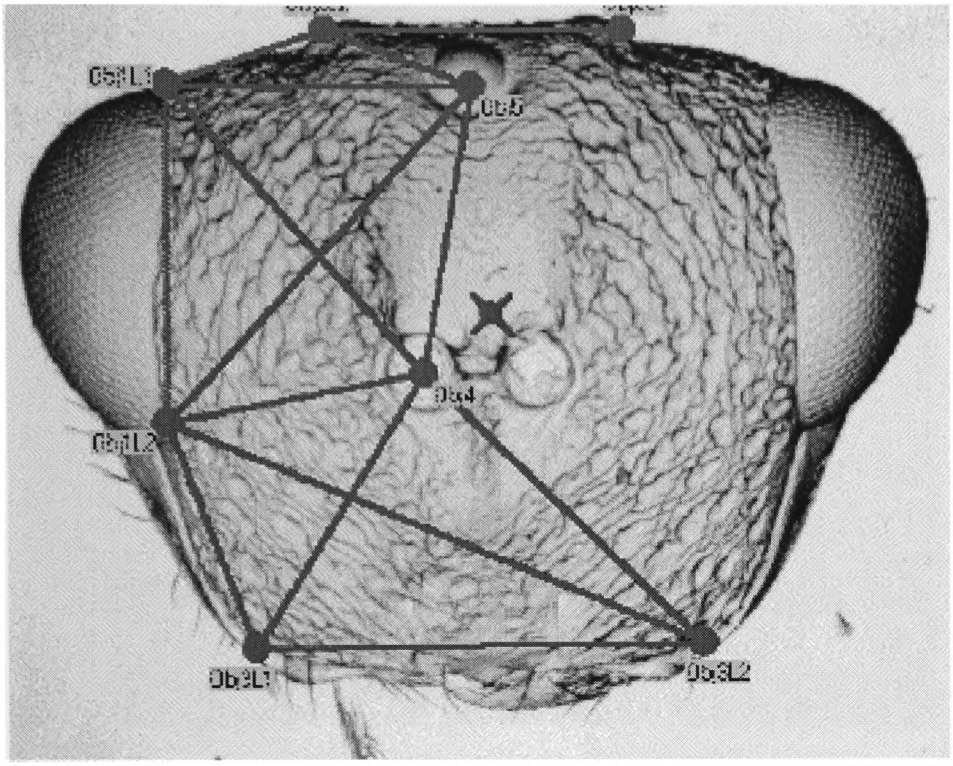
e301


Obj2L2
Obj2L1
Obj1L1
Obj5
Obj4
Obj1L2
Obj3L1
Obj3L2

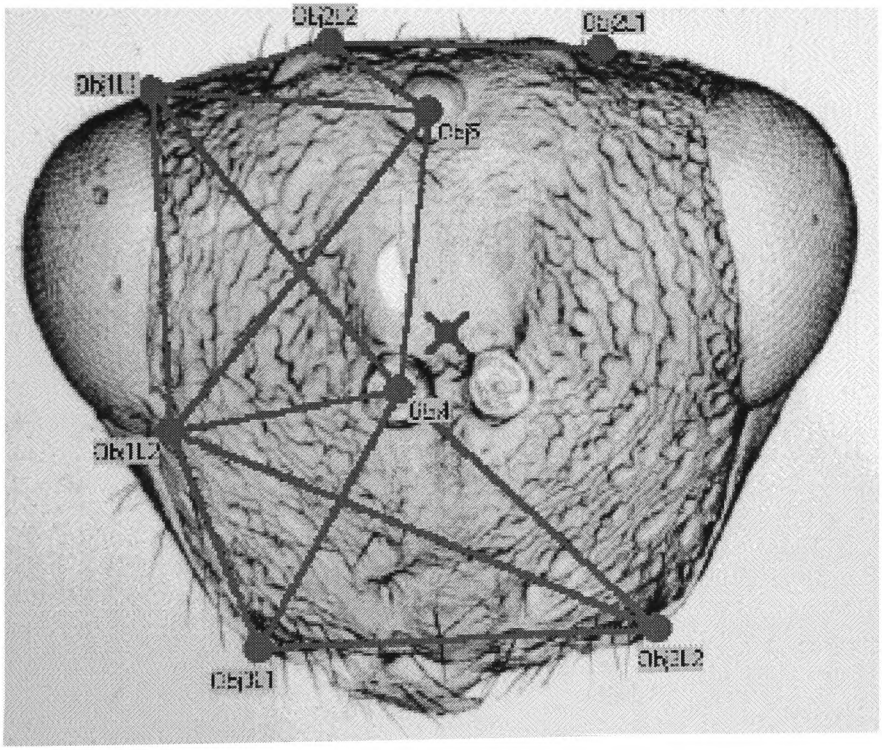
e302


Obj2L2
Obj2L1
Obj1L1
Obj5
Obj4
Obj1L2
Obj3L1
Obj3L2

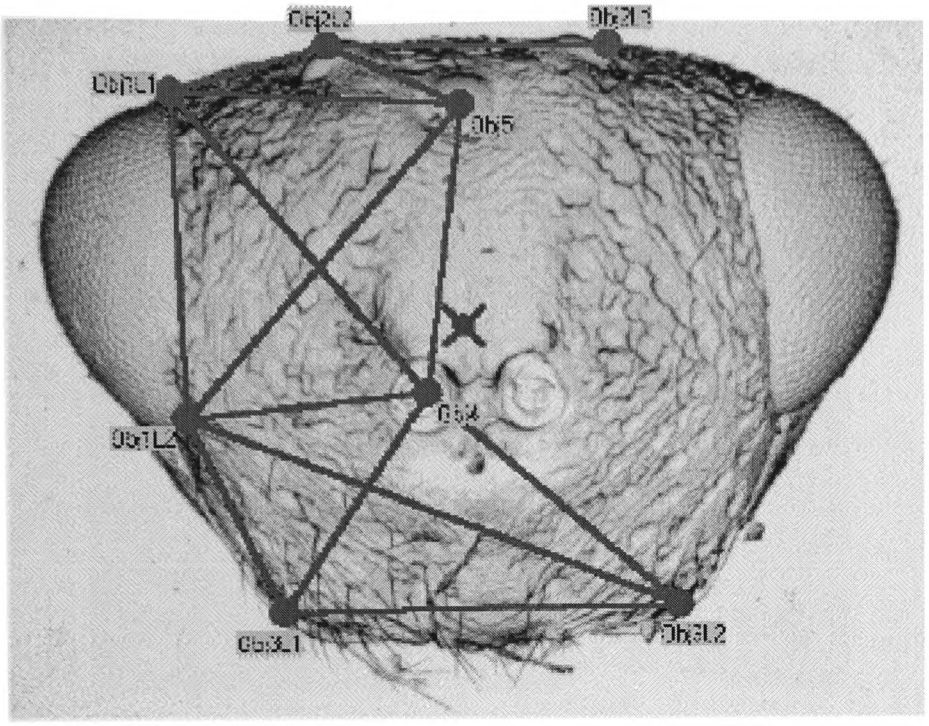
e303


Ob1L1
Ob5
Ob4
Ob1L2
Ob3L1
Ob3L2

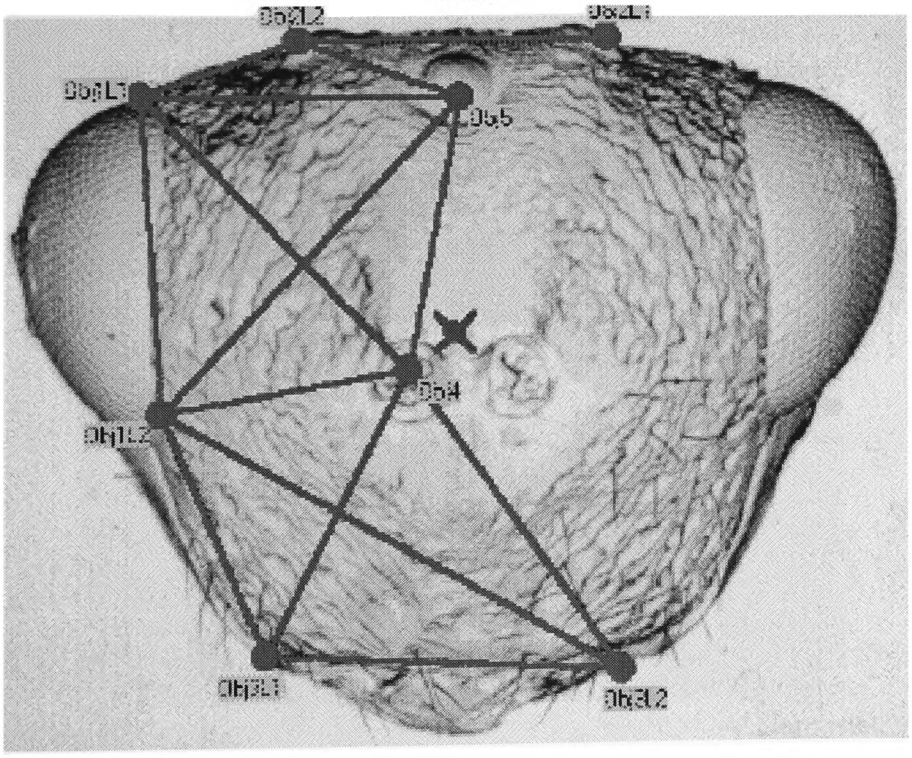
e304


Ob2L2
Ob2L1
Ob1L1
Ob5
Ob4
Ob1L2
Ob3L1
Ob3L2

e305

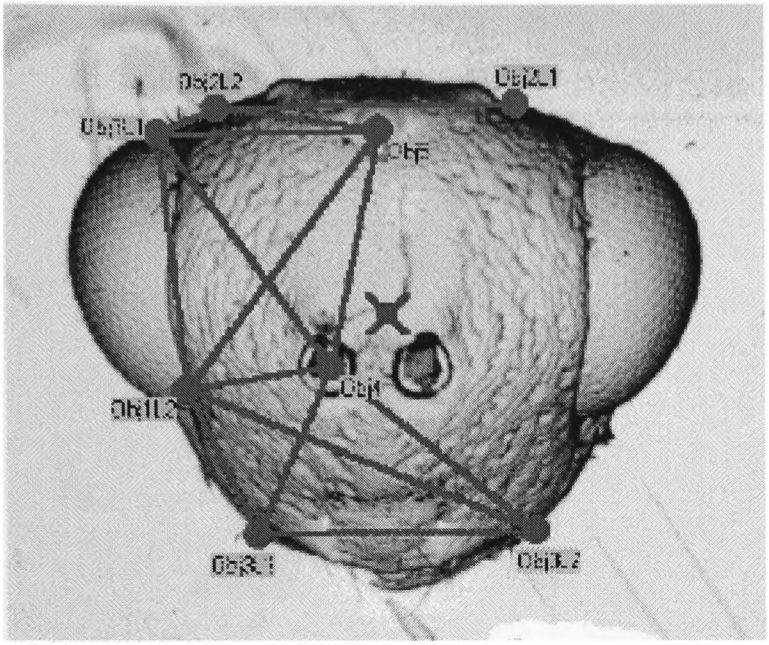

Obj2L2
Obj2L1
Obj1L1
Obj5
Obj4
Obj1L2
Obj3L1
Obj3L2

e306

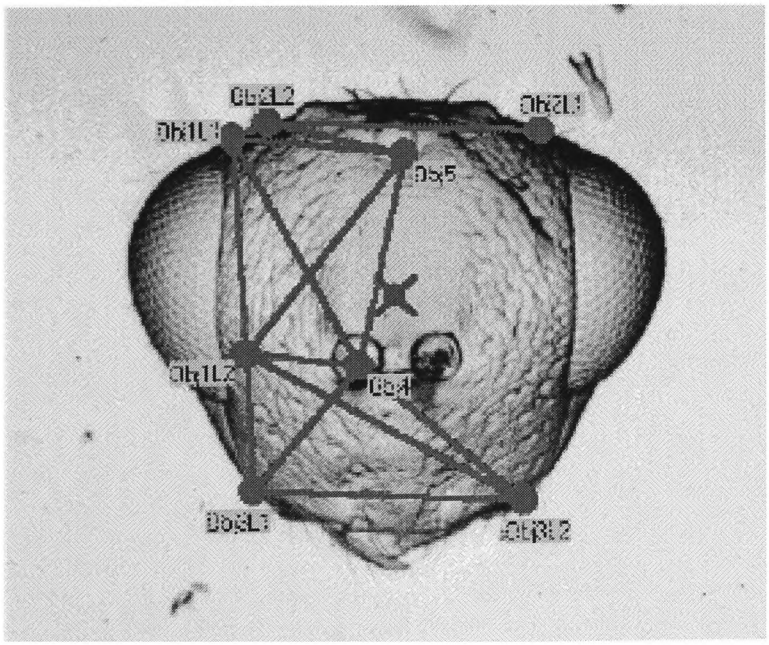

Obj2L2
Obj1L1
Obj1L2
Obj5
Obj4
Obj3L1
Obj3L2

e307

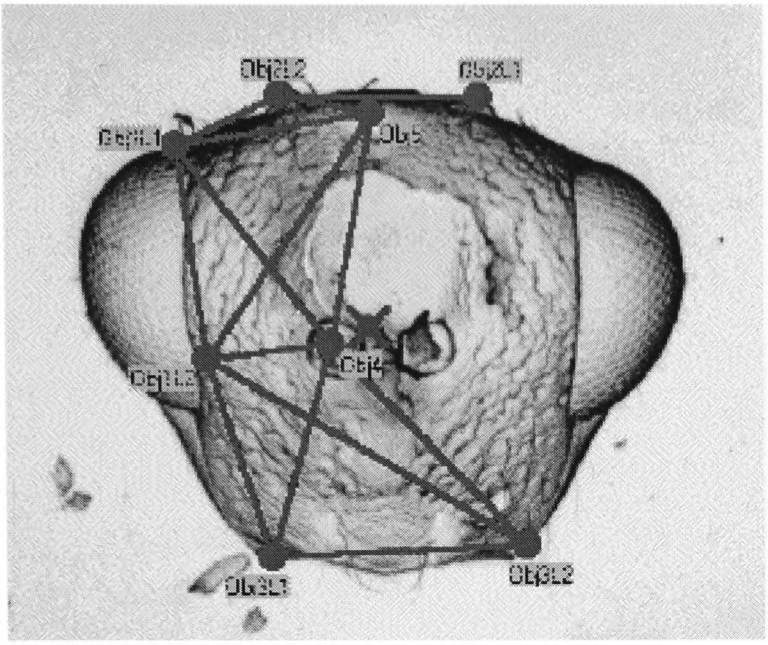
Species 4 : *Tetramesa fulvicollis*



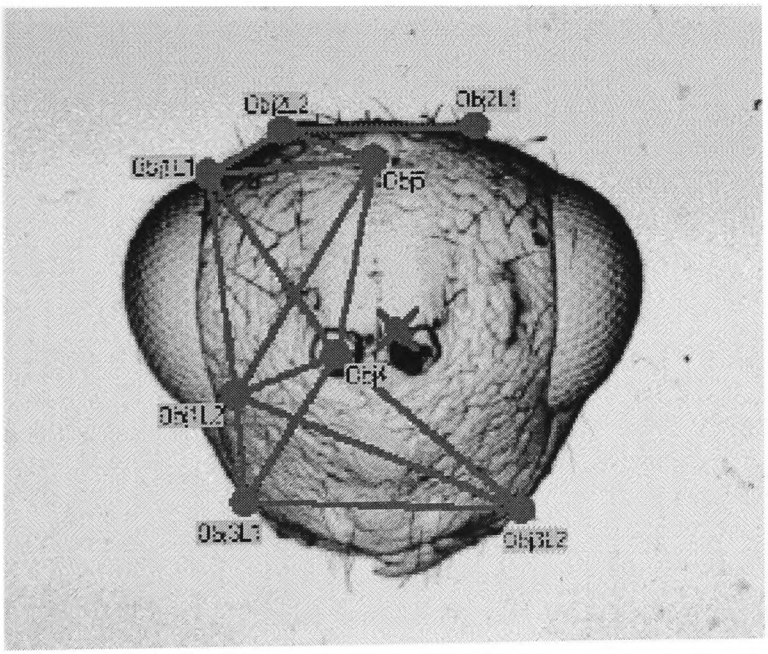
e400



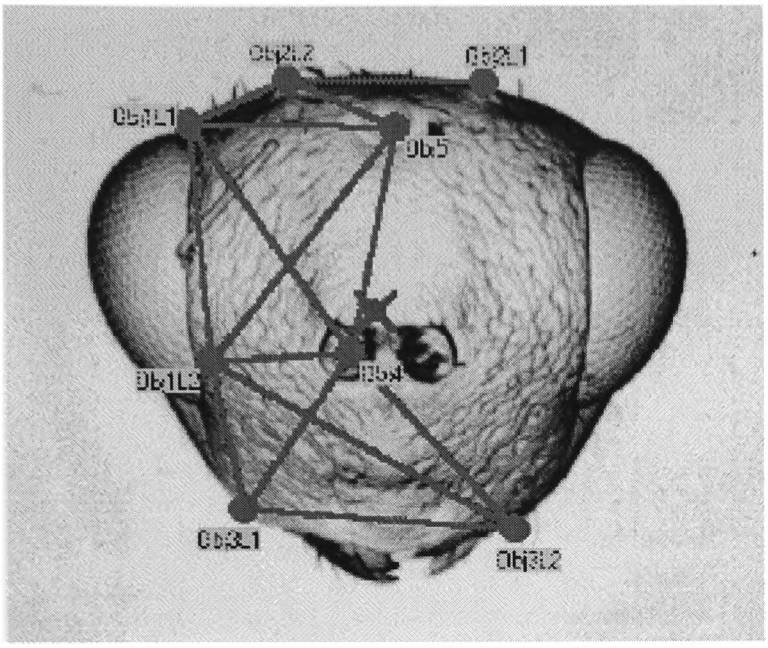
e401


Obj2L2
Obj2L1
Obj1L1
Obj
Obj4
Obj1L2
Obj3L3
Obj3L2

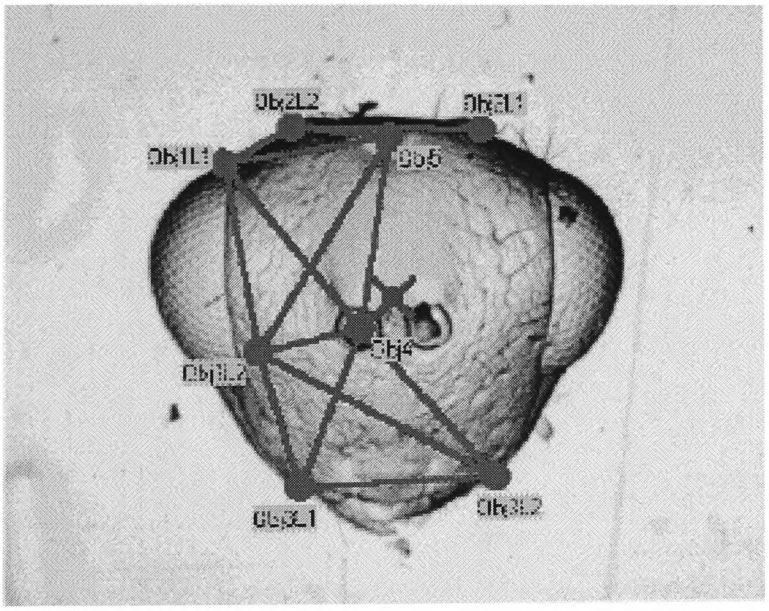
e402


Obj2L2
Obj2L1
Obj1L1
Obj
Obj4
Obj1L2
Obj3L3
Obj3L2

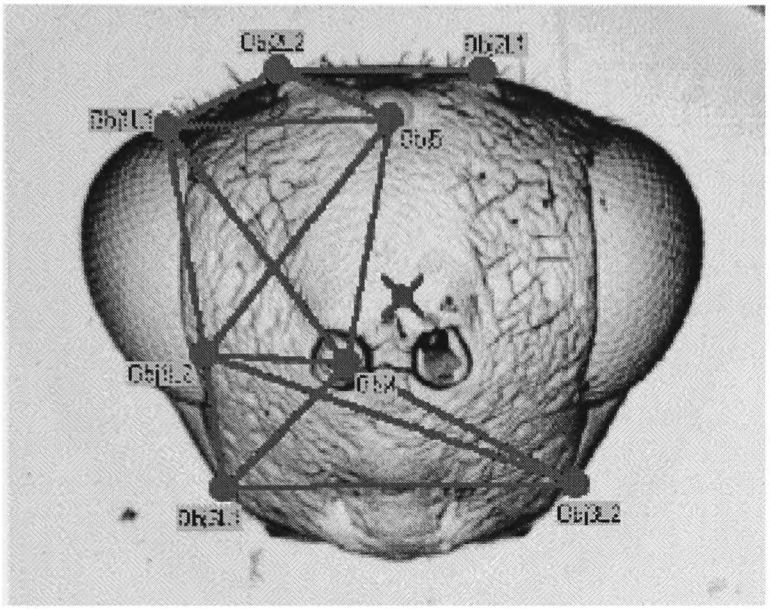
e403


Obj2L2
Obj2L1
Obj1L1
Obj5
Obj1L2
Obj4
Obj3L1
Obj3L2

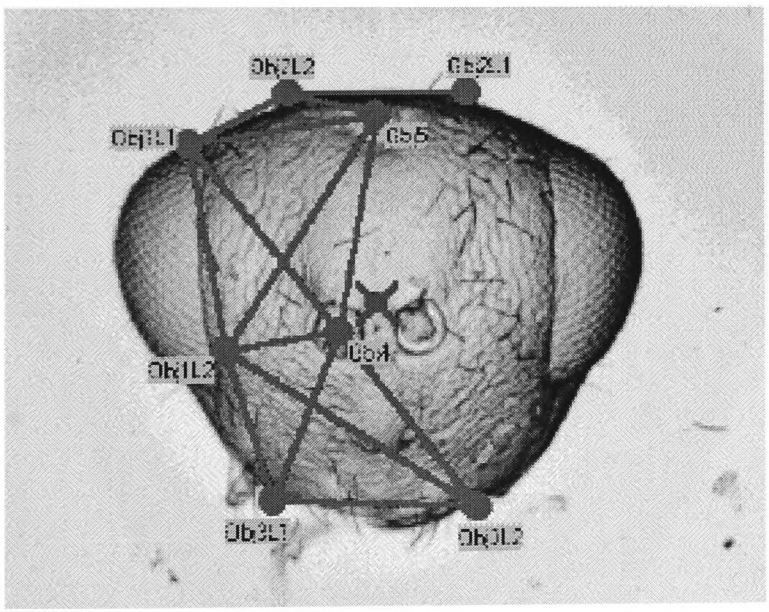
e404


Obj2L2
Obj2L1
Obj1L1
Obj5
Obj4
Obj3L2
Obj3L1
Obj3L2

e405

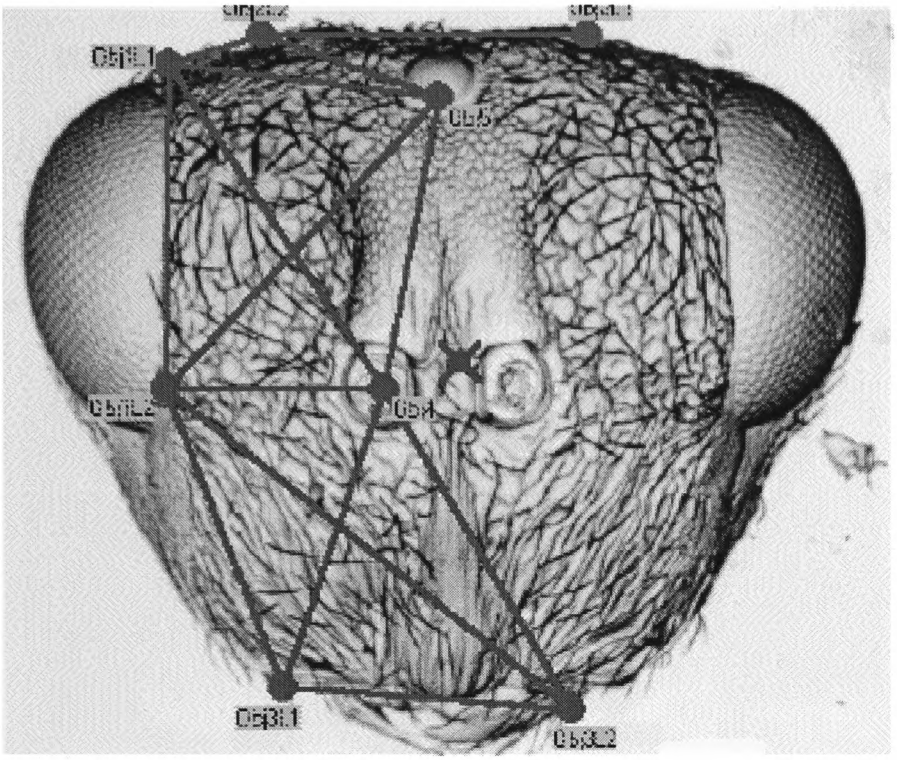


e406

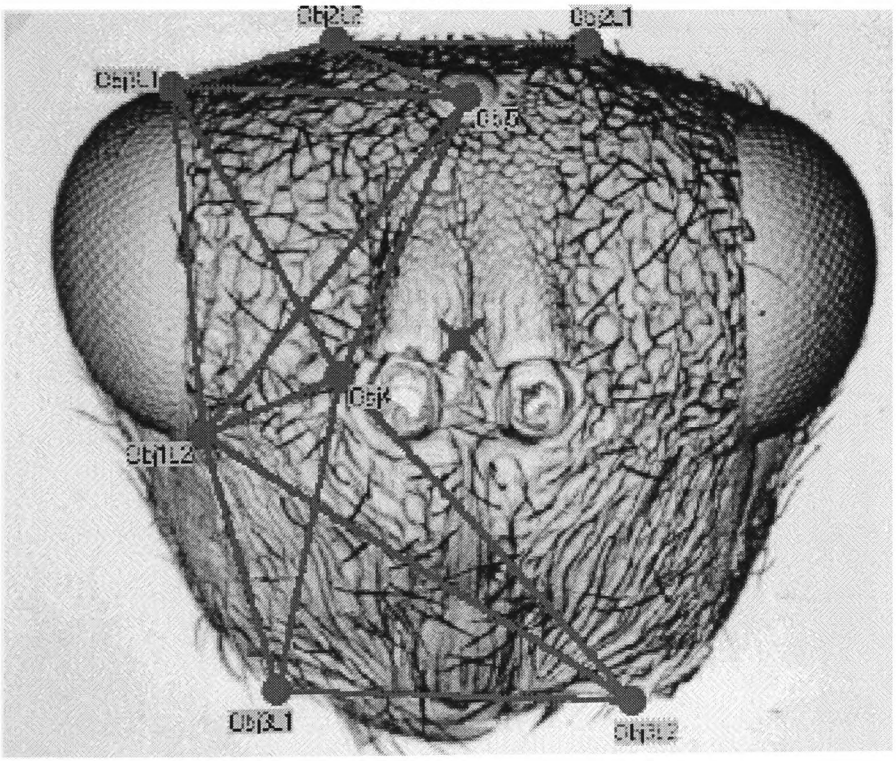


e407

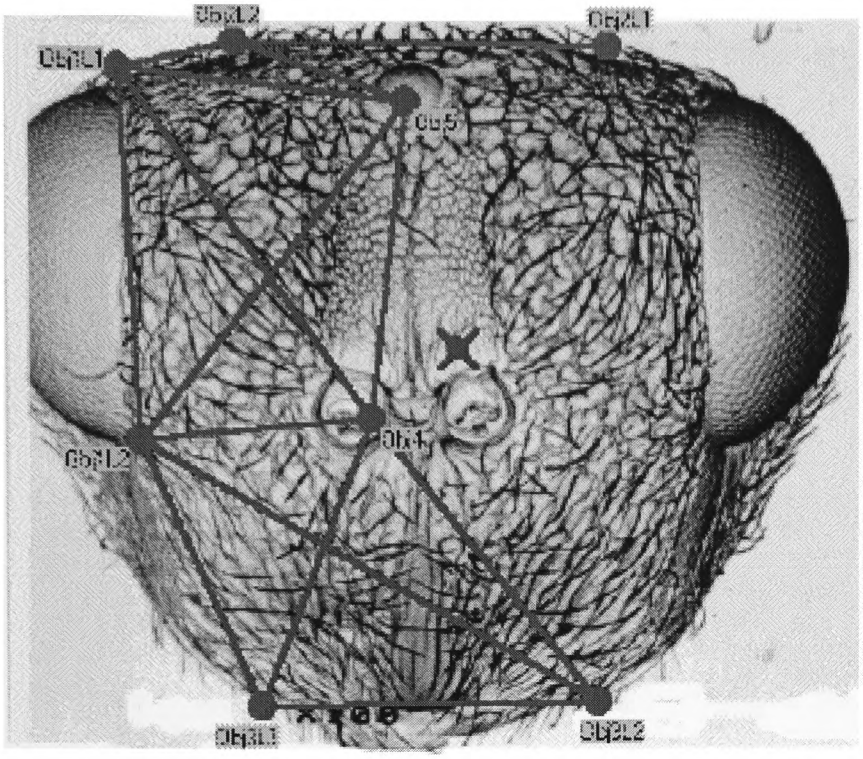
Species 5 : *Tetramesa eximia*



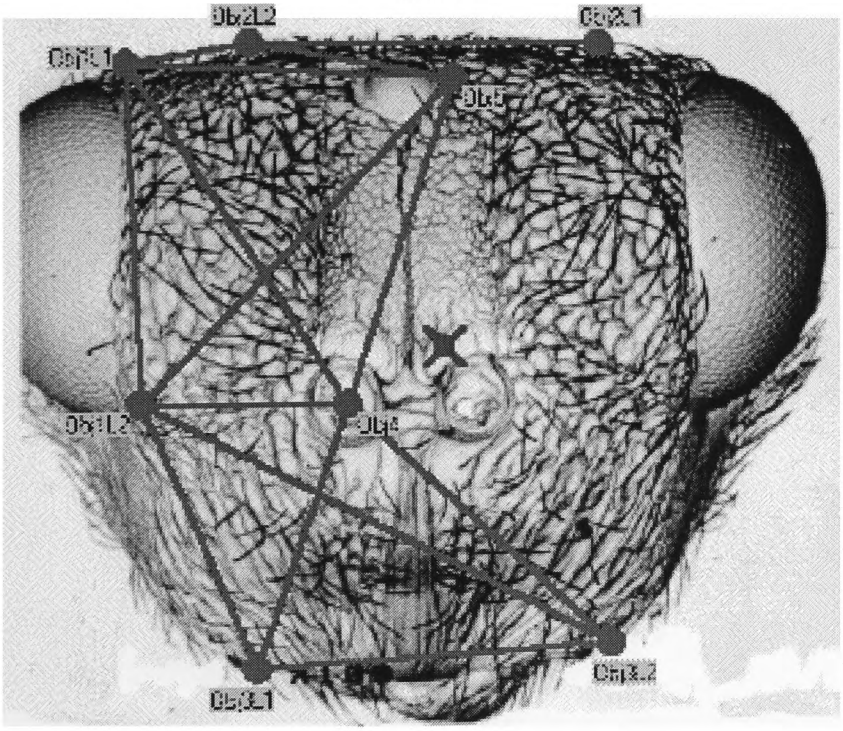
e500



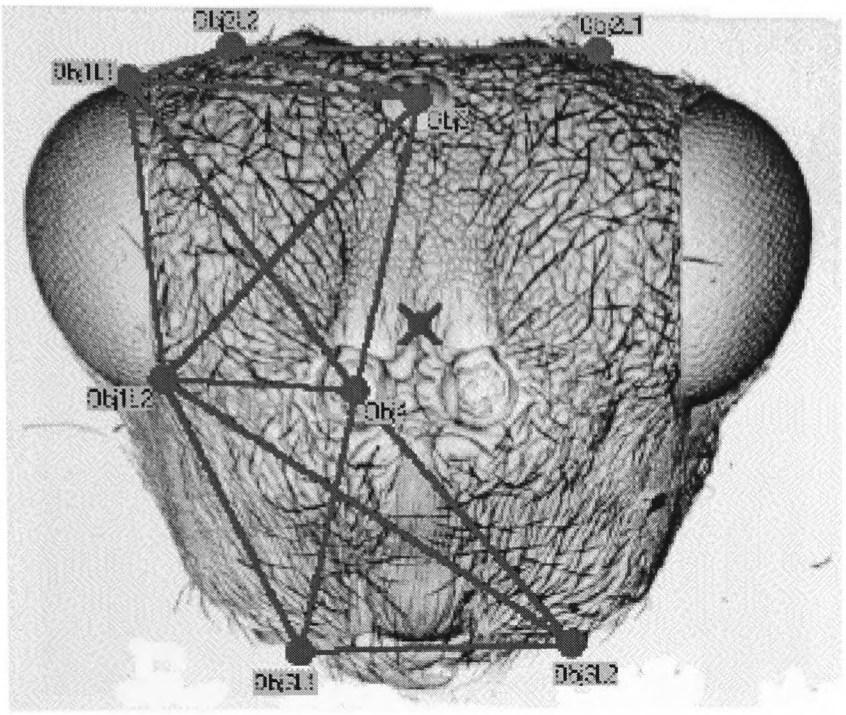
e501


Ob2L2
Ob2L1
Ob1L1
Ob5
Ob4
Ob1L2
Ob3L1
Ob3L2

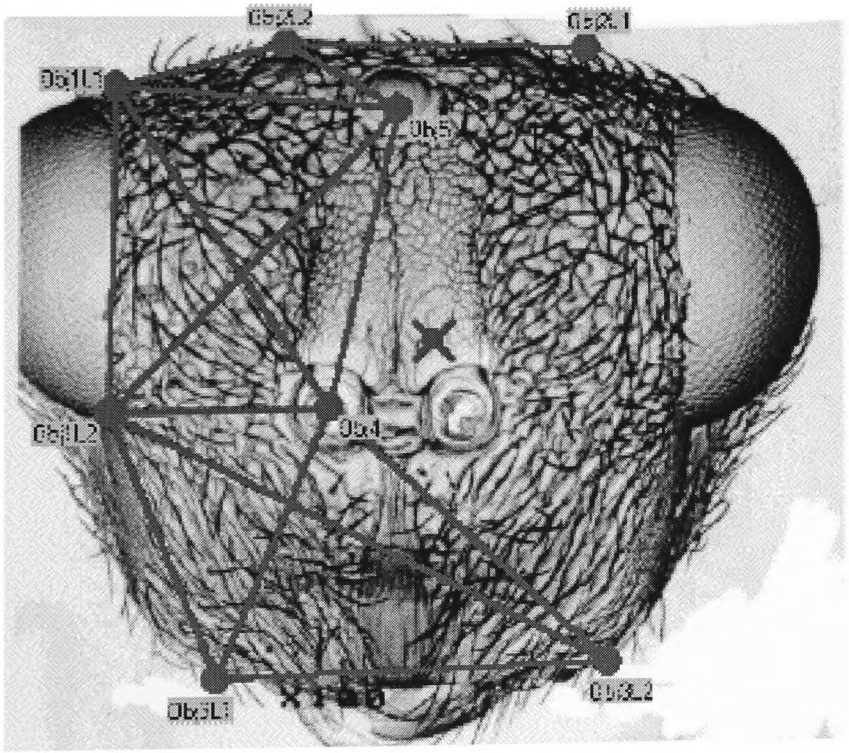
e502


Ob2L2
Ob2L1
Ob1L1
Ob5
Ob1L2
Ob4
Ob3L2
Ob3L1

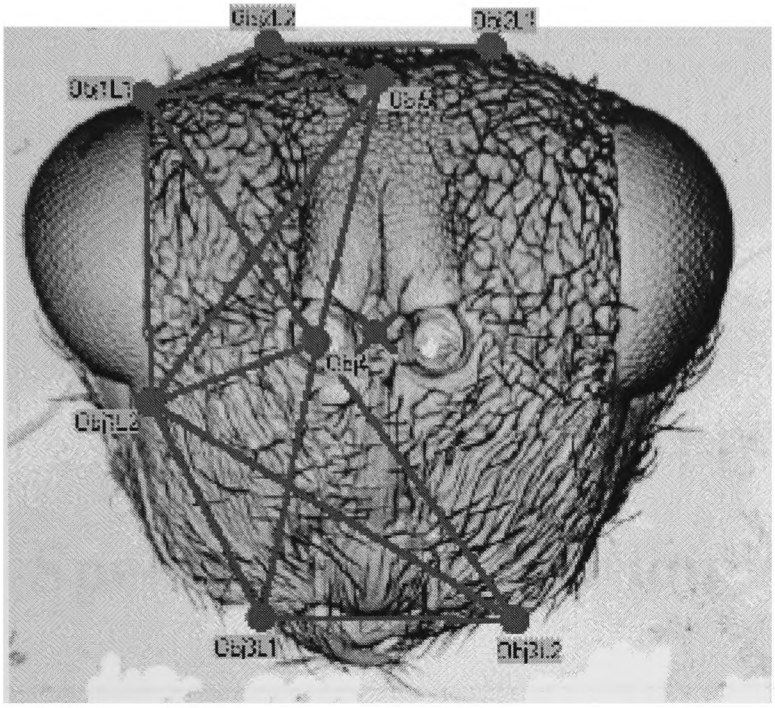
e503


Obj2L2
Obj2L1
Obj1L1
Obj5
Obj1L2
Obj4
Obj3L1
Obj3L2

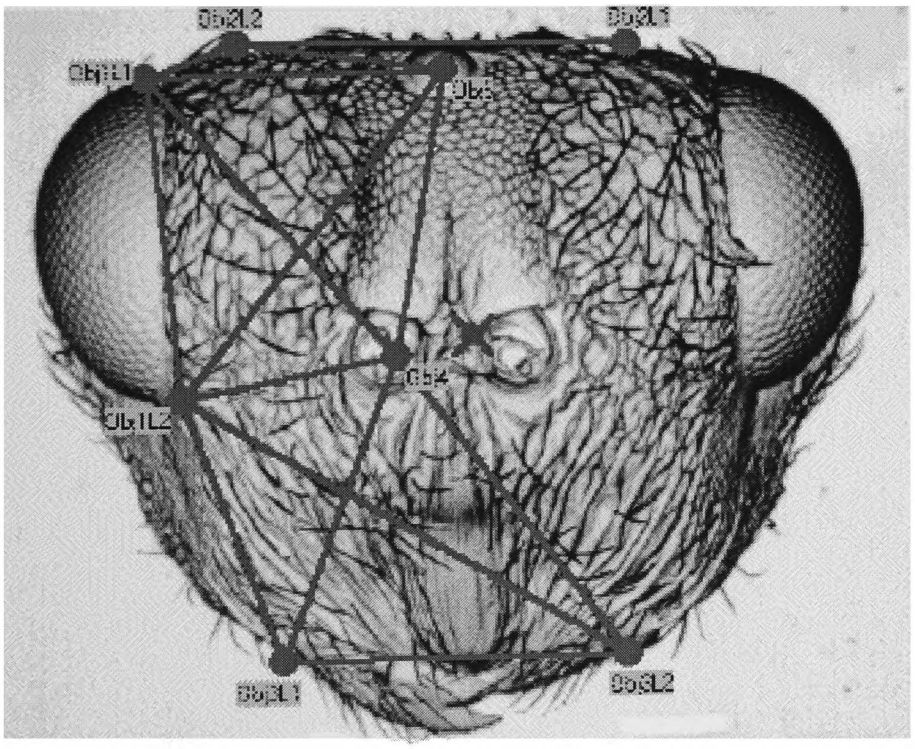
e504


Obj2L2
Obj2L1
Obj1L1
Obj5
Obj1L2
Obj4
Obj3L1
Obj3L2

e505


Obj2L2
Obj2L1
Obj1L1
Obj5
Obj4
Obj1L2
Obj3L1
Obj3L2

e506


Obj2L2
Obj2L1
Obj1L1
Obj5
Obj4
Obj1L2
Obj3L1
Obj3L2

e507

Appendix 9

Species Classification Results

Table 1. Species Classification Results for Experiment 1

	Species 1	Species 2	Species 3	Species 4	Species 5
e100	1.00	-1.00	-0.99	-1.00	-1.00
e101	1.00	-1.00	-0.98	-1.00	-1.00
e102	1.00	-1.00	-0.99	-1.00	-1.00
e103	0.97	-1.00	-1.00	-1.00	-1.00
e104	1.00	-1.00	-0.99	-1.00	-1.00
e105	0.98	-1.00	-0.98	-1.00	-1.00
e106	0.99	-1.00	-1.00	-1.00	-1.00
e107	0.99	-1.00	-1.00	-1.00	-1.00
e108	1.00	-1.00	-0.98	-1.00	-1.00
e109	0.98	-1.00	-1.00	-1.00	-1.00
e110	0.99	-1.00	-0.95	-1.00	-1.00
e111	0.93	-1.00	-0.48	-1.00	-1.00
e200	-1.00	1.00	-0.99	-1.00	-1.00
e201	-1.00	-1.00	0.68	-1.00	-0.99
e202	-1.00	0.97	-0.97	-1.00	-1.00
e203	-1.00	-1.00	0.95	-1.00	-1.00
e204	-1.00	1.00	-1.00	-1.00	-0.96
e205	-1.00	-1.00	-0.31	-1.00	0.66
e206	-0.97	-1.00	1.00	-1.00	-1.00
e300	-0.99	-1.00	0.96	-1.00	-0.96
e301	-1.00	0.88	-1.00	-0.97	0.99
e302	-1.00	-0.97	0.98	-1.00	-1.00
e303	-1.00	-0.73	-0.95	-1.00	-0.46
e304	-0.95	-1.00	0.99	-1.00	-0.99
e305	-0.98	-1.00	0.99	-1.00	-1.00
e306	-1.00	-1.00	-0.15	-1.00	0.21
e307	-1.00	-1.00	0.75	-1.00	-0.83
e400	-1.00	-1.00	-1.00	0.99	-0.96
e401	-1.00	-1.00	-1.00	1.00	-0.97
e402	-1.00	-1.00	-1.00	1.00	-0.96
e403	0.97	-1.00	-1.00	-0.97	-1.00
e404	-0.90	-1.00	-1.00	0.83	-1.00
e405	-0.99	-1.00	-1.00	0.99	-0.99
e406	-0.99	-1.00	-1.00	1.00	-0.98
e407	-1.00	-1.00	-1.00	0.99	-0.99
e500	-1.00	0.99	-1.00	-1.00	0.93
e501	0.55	-1.00	0.55	-1.00	-1.00
e502	-1.00	0.33	-1.00	-1.00	0.93
e503	-1.00	-0.98	-1.00	-0.19	1.00
e504	-1.00	0.99	-1.00	-0.98	0.97
e505	-1.00	-0.98	-1.00	-0.99	0.96
e506	-0.95	-1.00	0.31	-1.00	-1.00
e507	-1.00	-0.99	-0.97	-0.99	0.95

Table 2. Species Classification Results for Experiment 2

	Species 1	Species 2	Species 3	Species 4	Species 5
e100	1.00	-1.00	-0.99	-1.00	-1.00
e101	1.00	-1.00	-0.98	-1.00	-1.00
e102	1.00	-1.00	-1.00	-1.00	-1.00
e103	0.99	-1.00	-1.00	-1.00	-1.00
e104	1.00	-1.00	-1.00	-1.00	-1.00
e105	0.99	-1.00	-0.97	-1.00	-1.00
e106	1.00	-1.00	-1.00	-1.00	-1.00
e107	1.00	-1.00	-1.00	-0.99	-1.00
e108	1.00	-1.00	-0.97	-1.00	-1.00
e109	0.98	-1.00	-0.99	-1.00	-1.00
e110	0.99	-1.00	-0.95	-1.00	-1.00
e111	0.94	-1.00	-0.58	-1.00	-1.00
e200	-1.00	1.00	-0.99	-1.00	-1.00
e201	-1.00	-0.98	0.70	-1.00	-1.00
e202	-1.00	0.98	-0.96	-1.00	-0.99
e203	-1.00	-0.99	0.97	-1.00	-1.00
e204	-1.00	1.00	-1.00	-1.00	-0.98
e205	-1.00	-1.00	0.26	-1.00	0.56
e206	-0.99	-1.00	1.00	-1.00	-1.00
e300	-1.00	-1.00	0.95	-1.00	-0.97
e301	-1.00	0.62	-1.00	-0.99	0.99
e302	-1.00	-0.98	0.98	-1.00	-1.00
e303	-1.00	0.25	-0.88	-1.00	-0.53
e304	-0.92	-1.00	0.99	-1.00	-0.78
e305	-0.99	-1.00	1.00	-1.00	-0.99
e306	-1.00	-1.00	0.54	-1.00	0.64
e307	-1.00	-0.95	0.71	-1.00	-0.99
e400	-1.00	-1.00	-1.00	1.00	-0.97
e401	-1.00	-1.00	-1.00	1.00	-1.00
e402	-1.00	-1.00	-1.00	1.00	-0.96
e403	0.99	-1.00	-1.00	-0.94	-1.00
e404	-0.98	-1.00	-1.00	0.76	-1.00
e405	-1.00	-1.00	-1.00	1.00	-1.00
e406	-1.00	-1.00	-1.00	1.00	-1.00
e407	-1.00	-1.00	-1.00	0.99	-1.00
e500	-1.00	0.99	-1.00	-1.00	0.90
e501	0.82	-1.00	0.64	-1.00	-1.00
e502	-1.00	0.63	-0.99	-1.00	0.91
e503	-1.00	-0.96	-1.00	-0.42	1.00
e504	-1.00	0.85	-1.00	-0.94	0.99
e505	-1.00	-0.98	-0.99	-0.99	1.00
e506	-0.97	-1.00	-0.53	-1.00	-1.00
e507	-1.00	-1.00	-0.98	-1.00	0.96

Table 3. Species Classification Results for Experiment 3

	Species 1	Species 2	Species 3	Species 4	Species 5
e100	1.00	-1.00	-0.96	-1.00	-1.00
e101	1.00	-1.00	-0.99	-1.00	-1.00
e102	1.00	-1.00	-1.00	-1.00	-1.00
e103	0.96	-1.00	-1.00	-1.00	-0.91
e104	1.00	-1.00	-1.00	-1.00	-1.00
e105	0.98	-1.00	-1.00	-1.00	-1.00
e106	1.00	-1.00	-1.00	-0.99	-1.00
e107	0.98	-1.00	-1.00	-0.98	-0.97
e108	1.00	-1.00	-1.00	-1.00	-1.00
e109	-0.50	-0.99	-1.00	-1.00	-0.98
e110	1.00	-1.00	-1.00	-1.00	-1.00
e111	0.94	-1.00	-1.00	-1.00	-1.00
e200	-1.00	1.00	-0.96	-1.00	-1.00
e201	-1.00	0.98	-0.90	-1.00	-1.00
e202	-1.00	0.84	-0.96	-1.00	-1.00
e203	-1.00	-0.21	0.97	-1.00	-1.00
e204	-1.00	1.00	-1.00	-0.99	-1.00
e205	-1.00	-1.00	1.00	-1.00	-1.00
e206	-1.00	-0.98	0.98	-1.00	-1.00
e300	-1.00	-1.00	0.98	-1.00	-1.00
e301	-1.00	-1.00	0.95	-0.98	-1.00
e302	-1.00	-0.94	0.94	-1.00	-1.00
e303	-1.00	-0.99	0.99	-1.00	-1.00
e304	-1.00	-1.00	1.00	-1.00	-1.00
e305	-1.00	-0.96	0.99	-1.00	-1.00
e306	-1.00	-1.00	1.00	-0.90	-1.00
e307	-1.00	-1.00	0.98	-1.00	-1.00
e400	-1.00	-1.00	-1.00	0.99	-0.99
e401	-1.00	-1.00	-1.00	1.00	-0.98
e402	-1.00	-0.99	-1.00	0.15	-0.32
e403	-0.78	-1.00	-1.00	0.45	-1.00
e404	-1.00	-1.00	-1.00	0.90	-1.00
e405	-1.00	-1.00	-1.00	0.97	-0.99
e406	-1.00	-1.00	-0.97	1.00	-1.00
e407	-1.00	-0.99	-1.00	0.87	-0.99
e500	-1.00	-0.93	-1.00	-1.00	0.98
e501	-0.99	-0.99	-1.00	-1.00	0.97
e502	-1.00	-0.92	-0.99	-1.00	0.62
e503	-1.00	-1.00	-1.00	-0.98	0.99
e504	-1.00	-0.91	-1.00	-0.96	0.92
e505	-1.00	-0.97	-0.46	-1.00	-0.96
e506	-1.00	0.41	-1.00	-1.00	-0.98
e507	-1.00	-1.00	-1.00	-1.00	0.99

Table 4. Species Classification Results for Experiment 4

	Species 1	Species 2	Species 3	Species 4	Species 5
e100	1.00	-1.00	-0.95	-1.00	-1.00
e101	1.00	-1.00	-1.00	-1.00	-1.00
e102	1.00	-1.00	-1.00	-1.00	-1.00
e103	0.99	-1.00	-1.00	-0.99	-0.99
e104	1.00	-1.00	-1.00	-1.00	-1.00
e105	0.99	-1.00	-1.00	-1.00	-1.00
e106	1.00	-1.00	-1.00	-0.99	-1.00
e107	0.97	-1.00	-0.98	-0.98	-1.00
e108	1.00	-1.00	-1.00	-1.00	-1.00
e109	0.72	-1.00	-1.00	-0.99	-0.47
e110	1.00	-1.00	-1.00	-1.00	-1.00
e111	0.99	-1.00	-1.00	-1.00	-1.00
e200	-1.00	0.96	-0.91	-1.00	-1.00
e201	-1.00	0.90	-0.89	-1.00	-1.00
e202	-1.00	0.84	-0.89	-1.00	-0.98
e203	-1.00	-0.48	0.85	-1.00	-1.00
e204	-1.00	1.00	-1.00	-1.00	-1.00
e205	-1.00	-1.00	1.00	-1.00	-1.00
e206	-1.00	-0.99	0.97	-1.00	-1.00
e300	-1.00	-1.00	0.96	-1.00	-1.00
e301	-1.00	-0.96	0.93	-0.99	-1.00
e302	-1.00	-0.91	0.91	-1.00	-1.00
e303	-1.00	-0.92	0.94	-1.00	-1.00
e304	-1.00	-1.00	1.00	-1.00	-1.00
e305	-1.00	-0.95	0.97	-1.00	-1.00
e306	-1.00	-1.00	1.00	-0.99	-1.00
e307	-1.00	-1.00	0.96	-1.00	-1.00
e400	-1.00	-1.00	-1.00	0.97	-0.97
e401	-1.00	-1.00	-1.00	1.00	-1.00
e402	-1.00	-0.99	-1.00	0.94	0.64
e403	-0.15	-1.00	-1.00	0.71	-1.00
e404	-1.00	-1.00	-1.00	0.99	-1.00
e405	-0.98	-1.00	-1.00	1.00	-1.00
e406	-1.00	-1.00	-0.99	0.99	-1.00
e407	-1.00	-0.99	-1.00	0.99	-0.99
e500	-1.00	-0.92	-1.00	-1.00	1.00
e501	-0.98	-1.00	-1.00	-1.00	1.00
e502	-1.00	-0.67	-0.94	-1.00	0.56
e503	-1.00	-1.00	-1.00	-0.98	1.00
e504	-1.00	-0.87	-1.00	-0.81	1.00
e505	-1.00	-0.71	0.07	-1.00	-0.96
e506	-0.99	-0.89	-1.00	-1.00	0.62
e507	-1.00	-1.00	-0.99	-1.00	0.98

Table 5. Species Classification Results for Experiment 5

	Species 1	Species 2	Species 3	Species 4	Species 5
e100	1.00	-1.00	-0.94	-1.00	-1.00
e101	1.00	-1.00	-0.95	-1.00	-1.00
e102	1.00	-1.00	-1.00	-1.00	-0.99
e103	0.99	-1.00	-1.00	-1.00	-0.75
e104	1.00	-1.00	-1.00	-1.00	-1.00
e105	1.00	-1.00	-1.00	-1.00	-1.00
e106	1.00	-1.00	-1.00	-1.00	-1.00
e107	0.97	-1.00	-1.00	-0.97	-1.00
e108	1.00	-1.00	-1.00	-1.00	-1.00
e109	0.54	-1.00	-1.00	-0.97	-0.47
e110	1.00	-1.00	-1.00	-1.00	-1.00
e111	0.99	-1.00	-1.00	-1.00	-1.00
e200	-1.00	0.99	-0.94	-1.00	-1.00
e201	-1.00	0.97	-0.86	-1.00	-1.00
e202	-1.00	0.90	-0.90	-1.00	-0.96
e203	-1.00	-0.16	0.94	-1.00	-1.00
e204	-1.00	1.00	-1.00	-0.99	-1.00
e205	-1.00	-1.00	1.00	-1.00	-1.00
e206	-1.00	-0.99	0.97	-1.00	-1.00
e300	-1.00	-1.00	0.94	-1.00	-1.00
e301	-1.00	-1.00	0.89	-0.98	-1.00
e302	-1.00	-0.94	0.86	-1.00	-1.00
e303	-1.00	-0.93	0.96	-1.00	-1.00
e304	-1.00	-1.00	1.00	-1.00	-1.00
e305	-1.00	-0.93	0.99	-1.00	-1.00
e306	-1.00	-1.00	1.00	-0.98	-1.00
e307	-1.00	-1.00	0.95	-1.00	-1.00
e400	-1.00	-1.00	-1.00	0.98	-0.98
e401	-1.00	-1.00	-1.00	1.00	-0.99
e402	-1.00	-1.00	-1.00	0.98	0.68
e403	0.21	-1.00	-1.00	0.69	-1.00
e404	-1.00	-1.00	-1.00	0.95	-1.00
e405	-1.00	-1.00	-1.00	1.00	-0.97
e406	-1.00	-1.00	-0.98	0.98	-1.00
e407	-1.00	-1.00	-1.00	0.99	-0.94
e500	-1.00	-0.94	-1.00	-1.00	1.00
e501	-0.99	-0.99	-1.00	-1.00	1.00
e502	-1.00	-0.71	-0.99	-1.00	0.90
e503	-1.00	-1.00	-1.00	-0.98	1.00
e504	-1.00	-0.99	-1.00	-0.46	1.00
e505	-1.00	-0.83	-0.31	-1.00	-0.94
e506	-1.00	0.44	-1.00	-1.00	-0.31
e507	-1.00	-1.00	-1.00	-1.00	0.99

Table 6. Distances between species 2 and species 3 within wasp head feature space

	e300	e301	e302	e303	e304	e305	e306	e307	Mean distance
e200	0.47	0.34	0.63	0.36	0.52	0.38	0.51	0.49	0.46
e201	0.23	0.35	0.42	0.24	0.37	0.30	0.36	0.29	0.32
e202	0.40	0.59	0.19	0.39	0.49	0.55	0.52	0.37	0.44
e203	0.35	0.32	0.55	0.30	0.38	0.16	0.31	0.43	0.35
e204	0.33	0.48	0.32	0.32	0.46	0.49	0.47	0.30	0.40
e205	0.30	0.20	0.60	0.25	0.26	0.32	0.26	0.34	0.32
e206	0.21	0.38	0.31	0.19	0.25	0.26	0.31	0.26	0.27
									2.55
							Species distance =		0.36

Table 7. Distances between replicants within species 2

	e200	e201	e202	e203	e204	e205	e206	Mean distance
e200	0.00	0.38	0.67	0.30	0.54	0.37	0.42	0.38
e201	0.38	0.00	0.43	0.27	0.30	0.35	0.25	0.28
e202	0.67	0.43	0.00	0.60	0.26	0.60	0.38	0.42
e203	0.30	0.27	0.60	0.00	0.51	0.28	0.31	0.32
e204	0.54	0.30	0.26	0.51	0.00	0.51	0.34	0.35
e205	0.37	0.35	0.60	0.28	0.51	0.00	0.35	0.35
e206	0.42	0.25	0.38	0.31	0.34	0.35	0.00	0.29
								2.41
							Species distance =	0.34

Table 8. Distances between species 1 and species 3 within wasp head feature space

	e300	e301	e302	e303	e304	e305	e306	e307	Mean distance
e100	0.57	0.79	0.71	0.68	0.56	0.68	0.68	0.60	0.75
e101	0.70	0.98	0.68	0.81	0.72	0.83	0.83	0.71	0.89
e102	0.77	1.04	0.76	0.88	0.79	0.91	0.88	0.78	0.97
e103	0.66	0.92	0.62	0.75	0.72	0.86	0.83	0.64	0.86
e104	0.77	1.04	0.80	0.89	0.78	0.91	0.89	0.79	0.98
e105	0.68	0.95	0.57	0.76	0.74	0.85	0.83	0.66	0.86
e106	0.72	0.98	0.80	0.85	0.75	0.88	0.84	0.74	0.94
e107	0.71	0.89	0.86	0.81	0.68	0.80	0.77	0.76	0.90
e108	0.77	1.04	0.67	0.86	0.83	0.88	0.91	0.79	0.96
e109	0.65	0.88	0.57	0.72	0.74	0.77	0.80	0.67	0.83
e110	0.76	1.04	0.63	0.85	0.82	0.90	0.92	0.76	0.95
e111	0.59	0.86	0.49	0.68	0.66	0.75	0.74	0.57	0.76
									6.21
								Species distance =	0.89

Table 9. Distances between species 5 and species 3 within wasp head feature space

	e300	e301	e302	e303	e304	e305	e306	e307	Mean distance
e500	0.41	0.47	0.63	0.44	0.57	0.61	0.62	0.39	0.52
e501	0.52	0.70	0.55	0.57	0.60	0.62	0.69	0.56	0.60
e502	0.29	0.39	0.48	0.30	0.39	0.41	0.40	0.30	0.37
e503	0.46	0.46	0.65	0.45	0.48	0.60	0.49	0.48	0.51
e504	0.45	0.45	0.58	0.42	0.56	0.60	0.55	0.43	0.50
e505	0.36	0.23	0.56	0.25	0.37	0.30	0.35	0.43	0.36
e506	0.45	0.68	0.37	0.51	0.57	0.63	0.64	0.45	0.54
e507	0.34	0.56	0.49	0.43	0.42	0.60	0.51	0.30	0.45
									3.85
								Species distance =	0.48

Appendix 10

Papers

Papers

Angel, P.N., Morris, C.W. "A Neural Network Pipeline Architecture for Image Feature Extraction". *1st Int. Conf. on Imaging Science, Systems and Technology*. ed. Arabnia, H.R. pp. 311-318. 1997.

Angel, P.N., Morris, C.W. "Analysing the Mallat Wavelet Transform to Delineate Contour and Textural Features". Submitted to *Computer Vision and Image Understanding*. Awaiting Response.

A Neural Network Pipeline Architecture for Image Feature Extraction

Mr. Paul Angel, Mr. Colin Morris
Department of Computer Studies
University of Glamorgan
Pontypridd, Wales, UK

Abstract - Neural network architectures for image feature extraction use adaptive layers at the feature extraction level to learn about local image features in order to eliminate a combinatorial explosion at the classification stage which can result from geometric transformations. Training such networks limits their application to the chosen domain. A new neural network architecture is presented consisting of a pipeline of neural layers which extract high level features from an image and resolve translation variance. This has been used to extract features from images of biological specimens. From the extracted features a region hierarchy can be derived and boundaries are reduced to a set of feature points, or boundary primitives which exhibit rotation and scale invariance. Such feature descriptors are adequate for geometric shapes, but are not invariant to deformation present in images of biological specimens.

Keywords: feature extraction, neural networks, segmentation, morphometrics

1 Introduction

Artificial neural networks are used in image feature extraction and subsequent pattern recognition because of their ability to handle noisy data and their inherent parallelism [7]. Traditionally, image understanding consists of preprocessing, feature extraction and classification which incorporates knowledge and processes to reason about the extracted features [9].

Generally, there are two approaches to image feature handling. The first involves extracting features from an image before passing them to a classifier which must resolve any variance resulting from geometric transformations. The second approach normalises the input features at the feature extraction stage before being passed to the classifier. This overcomes the combinatorial explosion of features resulting from geometric transformations [7]. Various neural network architectures have been developed based on this approach. The Neocognitron, a multi-layered network consisting of fixed and adaptive weights, has been used successfully to learn and recognise numeric characters invariant to translation [4]. Lin and Wang [8] propose a multi-layered architecture consisting of a position normalisation layer, a rotation normalisation layer and a feature extraction layer, the output of which is normalised for scale invariance. Widrow *et al* [14] propose an architecture based on multiple layers of ADALINES [13], and consists of an invariance net and a trainable classifier. ART networks have been used because of their stability and plasticity properties. Srinivasa and Jouaneh [11] have used the ART1 architecture for the recognition of untrained patterns, with a later extension to include a third layer, based on the principal of invariance, to reduce the number of nodes [12]. Zheng [16] uses the ART architecture in a two stage, grouping and description approach to image feature extraction and segmentation. Bhandarkar *et al* [3] also use neural networks for image segmentation. An extension to Kohonen's SOM [6], called the HSOFM (Hierarchical Self Organising Feature Map) is

used to build an abstraction tree of the regions in a given image. This overcomes the limitations of the SOM in that the number of regions does not have to be known a-priori.

This paper introduces an artificial neural network architecture comprising a pipeline of neural layers, each performing a specific function to extract features from an input image and output a hierarchical representation of the regions and edges that comprise the image. Unlike other multilayered architectures the interconnections between each layer are fixed, and serve more a communication rather than memory function, so no learning occurs at the feature extraction level. Instead, the pipeline of neural layers serves to reduce the data from an input image to a set of high level features, so a combinatorial explosion resulting from geometric transformations does not occur. Translational variance is resolved within the actual pipeline since an image is distributed across the nodes of each layer which work collectively to extract the local image structure. Section 2 of the paper will introduce the overall architecture, section 3 looks at how image features stored in the pipeline can be used to derive object descriptions which are invariant to rotation and scaling, section 4 shows experimental results and section 5 concludes with a summary of work carried out and future work that is planned.

2 Architecture

The neural network architecture consists of a pipeline of neural layers, each designed to extract particular high-level features from an input image. Its purpose is to reduce the data from an image into a set of high-level features for subsequent classification, so no learning is necessary. This eliminates the combinatorial explosion that might result from geometric transformations if local image features are used directly in the classification stage. There are 6 layers in the pipeline. The input layer, referred to as V1 simply because it relates to an early stage of visual processing [15], consists of an n*m array of nodes which correspond with the pixels in an input image. This layer sends the image information to the subsequent layers which carry out specific processing. The H and V layers store edge map information for the image with horizontal edges, or image state changes stored as active nodes in the H layer and vertical edges stored in the V layer. The RL layer uses the data in V1 to segment the image by assigning a unique label to each region. The edge map and segmentation data are passed to the BC layer which links edge points to form closed curves around each region. The last layer in the pipeline, BL, assigns a unique label to each boundary so they can be extracted for further processing. The feature extraction pipeline is shown in figure 1.

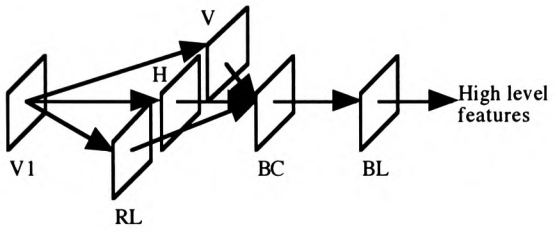


Figure 1. Neural network pipeline

2.1 Edge Map Representation

The H and V layers store horizontal and vertical state changes that occur between nodes in the V1 layer. Given that the V1 layer has n*m nodes, the H and V layers have n*m-1 and n-1*m nodes respectively. The nodes produce an output based on the inputs received from the V1 layer. A node in the H layer receives inputs from two horizontally neighbouring nodes in V1, $V1_{i,j}$ and $V1_{i,j+1}$, as shown in figure 2(a), where $0 \le i < n$ and $0 \le j < m-1$, and outputs a state which is based on the function OH_{ij} where

$$OH_{ij} = \begin{cases} 0 \text{ if } V1_{i,j} = V1_{i,j+1} \\ 1 \text{ otherwise} \end{cases} \quad (1)$$

Similarly for nodes in the V layer. These receive inputs from vertically neighbouring nodes in V1, $V1_{i,j}$ and $V1_{i+1,j}$, as shown in figure 2(b), where $0 \le i < n-1$ and $0 \le j < m$, and output a state based on the function OV_{ij} where

$$OV_{ij} = \begin{cases} 0 \text{ if } V1_{i,j} = V1_{i+1,j} \\ 1 \text{ otherwise} \end{cases} \quad (2)$$

The H and V layers form an edge map with excitory activations (+1) representing edge points.

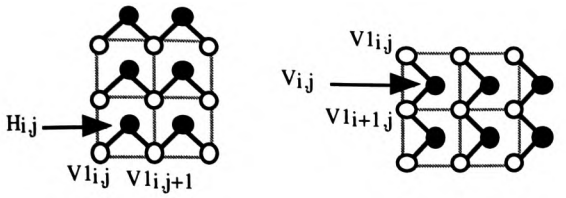


Figure 2. Interconnections between layers V1, H and V

2.2 Image Segmentation

The process of image segmentation can be reduced to one of vector quantisation [3], where an abstraction of the original image is formed showing the regions, or groups of connected pixels that satisfy some given homogeneity criteria. Within the neural network pipeline, segmentation is carried out in the RL (Region Labelling) layer. Here the nodes work collectively with their neighbours to form the different regions that comprise the image and because of the neural architecture this can be done in parallel which overcomes the problems of relabelling encountered in sequential algorithms [5]. An RL node is selected to represent each region which introduces the problem of selecting such nodes. Seeded Region Growing [1] requires a set of seed points to be set up, either through user interaction or through a selection algorithm incorporating high-level knowledge about the image. Vector quantisation requires that we know what regions will comprise an image [3] and Pyramidal methods require multiple layers to represent the image at different resolutions [3].

Since the RL layer is a single 2-dimensional array of nodes, a seeded region growing method is used in favour of the multi-dimensional approach. However, instead of having to provide selection criteria to select seed points, the RL layer assigns each node a unique strength. Given an arbitrary group of nodes that represents a region in an image, one node will have a greater strength than the others. This node can represent the region, whose strength, or label can be taken on by the connected nodes. The resulting segmentation remains a function of the given homogeneity criteria, but the image attributes of the pixel corresponding to the dominant node represents the characteristics of the region as a whole. The operation of each RL node can be made simple because of this organisation. Each node outputs its strength to its neighbouring nodes and receives their strengths as input. Given k inputs, I1..Ik, each RL node, RLij, updates its state according to the rule

$$\text{IF } \text{strength}_{i,j} < \text{MAX}(I_1,..,I_k) \quad \text{then } \text{strength}_{i,j} = \text{MAX}(I_1,..,I_k) \tag{3}$$

Each input Ii is a product of the output of a neighbouring RL node, Oxy, with some homogeneity function, H(nij,nxy) and is defined as

$$I_i = O_{xy}\, H(n_{ij},n_{xy}) \tag{4}$$

Here H(nij,nxy) is defined as the equality between the value of node V1ij and its neighbour V1xy.

$$H(V1_{ij},V1_{xy}) = \begin{cases} 1 \text{ if } V1_{ij} = V1_{xy} \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

The connectedness of the RL layer determines which neighbouring nodes are relevant. For a 4-connected RL layer, each node RLij receives input from and outputs to the neighbouring nodes RLi-1,j, RLi+1,j, RLi,j-1 and RLi,j+1. Given the homogeneity function in (5), the inputs to node RLij are defined as

$$I_1 = O_{i-1,j}\, H(V1_{ij},V1_{i-1,j}) \tag{6}$$

$$I_2 = O_{i+1,j}\, H(V1_{ij},V1_{i+1,j}) \tag{7}$$

$$I_3 = O_{i,j-1}\, H(V1_{ij},V1_{i,j-1}) \tag{8}$$

$$I_4 = O_{i,j+1}\, H(V1_{ij},V1_{i,j+1}) \tag{9}$$

For an 8-connected RL layer, diagonal neighbours RLi-1,j-1, RLi-1,j+1, RLi+1,j-1 and RLi+1,j+1 need to be considered. The inputs from these nodes, I5..I8 respectively, are defined as

$$I_5 = O_{i-1,j-1}\, H(V1_{ij},V1_{i-1,j-1}) \tag{10}$$

$$I_6 = O_{i-1,j+1}\, H(V1_{ij},V1_{i-1,j+1}) \tag{11}$$

$$I_7 = O_{i+1,j-1}\, H(V1_{ij},V1_{i+1,j-1}) \tag{12}$$

$$I_8 = O_{i+1,j+1}\, H(V1_{ij},V1_{i+1,j+1}) \tag{13}$$

The ambiguity resulting from a 4 or 8 connected segmentation procedure can be

overcome with a hexagonal tessellation [5]. Figure 3 illustrates the interconnections between nodes for a 6-connected RL layer, where figure 3(a) shows the interconnections to and from RL nodes which lie on row 0 and subsequent even numbered rows while figure 3(b) shows interconnections between RL nodes on odd numered rows.

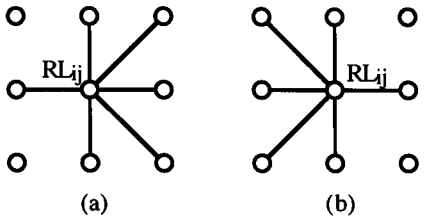


Figure 3. Interconnections for a 6-connected RL layer.

Nodes in the RL layer will continue to receive input and update their state according to equation (3) until the image has been segmented and the nodes in each region settle into equality. In order to determine this, each node is connected to a control node which receives an excitory signal (+1) if a node updates its state or an inhibitory signal (0) if its state remains constant. The output of the control node is defined as

$$C = \frac{\sum_{i}^{n} \sum_{j}^{m} U_{ij}}{nm} \qquad (14)$$

where Uij represents the update status of node RLij. When C=0, all nodes have settled into equality, that is all connected nodes in each region have the same strength, indicating that the segmentation process is complete.

2.3 Closed Curve formation

The BC (Boundary Creation) layer uses the label patterns stored in the RL layer to connect edge points stored as active nodes in the H and V layers forming closed curves around the regions that comprise an input image. Each region has an outer boundary and zero or more inner-boundaries surrounding regions nested within the outer region. An edge represents two boundaries, one for each of the two regions that meet. These relationships are used to derive a region hierarchy for the input image. Each BC node corresponds directly with nodes in the H and V layers and are labelled accordingly. This is shown in figure 4.

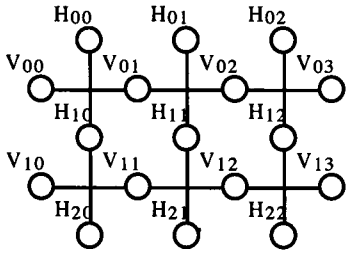


Figure 4. Relationships between nodes in the BC layer for a 3*4 region of the V1 layer.

Each BC node receives an activation signal (+1) from its corresponding H/V node. Inactive nodes in the H and V layers do not represent edge points so their corresponding BC nodes can remain dormant. Each BC node separates two regions, or two RL nodes and contains two boundary link sub-nodes. These are linked to the sub-nodes corresponding to the same region in one of the neighbouring BC nodes. These links are determined by the pattern of region labels input from the RL layer. For each BC node, two region label patterns are formed, one for each separated RL node. Region label patterns are ordered 4 digit binary strings based on the equality of a separated RL node with the relevant neighbouring RL nodes, where a 1 represents a label match and 0 a label mismatch. BC node Hij separates nodes RLij and RLi,j+1, and creates region label patterns using the inputs from RLi-1,j, RLi-1,j+1, RLi,j+1, RLi,j and RLi,j, RLi,j+1, RLi+1,j+1, RLi+1,j respectively. Table 1 shows which BC nodes are linked to from node Hij if the region label patterns based on nodes RLij and RLi,j+1 match those in the table. Similarly, BC node Vij separates nodes RLij and RLi+1,j, and creates region label patterns using the inputs from RLi,j, RLi,j+1, RLi+1,j+1, RLi+1,j and RLi,j-1, RLi,j, RLi+1,j, RLi+1,j-1 respectively.

Table 1. Region label patterns and resulting node links for BC node Hij.

Based on region node	Region label pattern	Create link to BC node
RL_{ij}	1001	$H_{i-1,j}$
RL_{ij}	0001	$V_{i-1,j}$
RL_{ij}	1101	$V_{i-1,j+1}$
$RL_{i,j+1}$	0110	$H_{i+1,j}$
$RL_{i,j+1}$	0111	$V_{i,j}$
$RL_{i,j+1}$	0100	$V_{i,j+1}$

Table 2 shows which BC nodes are linked to from node Vij if the region label patterns based on nodes RLij and RLi+1,j match those in the table.

Table 2. Region label patterns and resulting node links for BC node Vij.

Based on region node	Region label pattern	Create link to BC node
RL_{ij}	1100	$V_{i,j+1}$
RL_{ij}	1000	$H_{i,j}$
RL_{ij}	1110	$H_{i+1,j}$
$RL_{i+1,j}$	0011	$V_{i,j-1}$
$RL_{i+1,j}$	1011	$H_{i,j-1}$
$RL_{i+1,j}$	0010	$H_{i+1,j-1}$

If the RL nodes that make up a region label pattern are separated by 2 or 3 active BC nodes then the above patterns will create the appropriate links. If there are 4 active nodes separating two diagonally connected regions then the resulting region label pattern, 0101 or 1010, created for horizontal and vertical BC nodes respectively, will not be recognised. For these patterns, the links created are given in Table 3. However, if there are 4 active nodes where 4 regions converge, the appropriate links given in tables 1 and 2 will be made.

Table 3. BC node links created where 4 active RL nodes separate two image regions.

Based on region node	Region label pattern	Link from	Link to
RL_{ij}	0101	$H_{i,j}$	$V_{i-1,j+1}$
RL_{ij}	1010	$V_{i,j}$	$H_{i,j}$
$RL_{i,j+1}$	0101	$H_{i,j}$	$V_{i,j}$
$RL_{i,j+1}$	1010	$V_{i,j}$	$H_{i+1,j-1}$

Based on the region label patterns, the boundary link sub-nodes on an outer boundary are linked in a counter-clockwise direction, while those on an inner boundary are linked in a clockwise direction.

2.4 Boundary Labelling

The BL (Boundary Labelling) layer assigns each boundary a label so they can extracted from the neural network pipeline for further processing. Just as the nodes in the RL layer had a direct correspondence with nodes in the V1 layer, so nodes in the BL layer directly correspond with nodes in the BC layer. The labelling mechanism employed is similar to that described in section 2.2. Initially each boundary link sub-node is assigned a unique strength, so for any given boundary there will always be one dominant node. Each boundary link sub-node outputs its strength to its neighbouring sub-node, that is the node it is linked to in the BC layer. The updating of the nodes proceeds as given in equation (3). Since each boundary point links to only one neighbour, k=1, the problem of connectedness does not arise. The labelling process will continue until every node in each boundary settles into equality. This is determined by employing a control node, as described in section 2.2, for the BL layer.

3 Feature Representation

Once each boundary has been labelled, data in the neural network pipeline can be used to derive a suitable description of each object for subsequent classification. This description will depend upon the chosen application. However, one can now derive object descriptions from higher level image features and not low level pixel information obtained directly from the input image. This work looks at building object descriptions from a region and boundary hierarchy which is dynamically created based on the segmentation data in the RL layer and the intersection of inner and outer region boundaries stored in the BC and BL layers. Given an inner boundary, I1 for region R1, and an outer boundary O2 for nested region R2, if O2 $\supset$ I1 then R2 is represented as a child node of R1 in the region hierarchy. Inner and outer boundaries can be discriminated between because the outer boundary contains a greater number of points which are required to surround the region.

The neural network pipeline extracts features invariant to translation. Each boundary point is defined by its x,y position in image coordinates and can be normalised by subtracting the mean centre of the bounded region which is defined as

$$\overline{x}_j=\Sigma_{i=1}bx_{ij}/n \qquad \overline{y}_j=\Sigma_{i=1}by_{ij}/n$$

where bx,by represents the x and y position of

boundary point i in image coordinates and n equals the number of points in boundary j. In this normalised form, the centre of each boundary, or region is at the origin, so the relationship between two regions, Rij, can be maintained with

$$R_{ij} = \bar{x}_i - \bar{x}_j, \bar{y}_i - \bar{y}_j$$

Each boundary can be represented as a graph indicating the direction change of the boundary at each point, starting from the point furthest from the centre. Positive gradients represent convex features, negative gradients represent concave features while zero gradients represent straight edges. This representation can be used for different analyses, including Fourier and eigenshape analysis [10]. Here, the graph is reduced to a set of discrete feature points based on the gradient of each section of the graph. First, significant changes in the gradient are found to separate the individual features. Then the square of the section gradient and width are used to plot each section as a point in a 2-dimensional feature space. This space is partitioned into regions, as shown in figure 5 which determine the classification of each graph section. According to its location in this feature space, each graph section can be classified as a line, curve, point-corner (approximately 90 degrees), point (> 90 degrees), curve-corner (approximately 90 degrees), curve-turn (approximately 180 degrees) or a circle.

A point where the section-gradient^2(s)≤graph-gradient^2(g) is classified as a line (s < g/2) or curve (s ≥ g/2). The other points are classified by calculating the angle from the horizontal axis to the line which intersects the origin and the feature point. The boundary lines which partition the feature space are found through experimentation. The "circle" feature is a special case where the gradient of the feature graph remains constant since the direction change of each point on a circle remains constant. However distortions resulting from the discrete approximation of a circle requires the feature graph to be smoothed out. This is also required for "uneven" boundaries, where erroneous pixels can have a large effect on the shape of the feature graph. Once a graph section has been classified, a single point is chosen to represent that section, resulting in a more compact boundary representation. This representation can be made invariant to rotation by storing relative angles, calculated in a counter-clockwise direction, between each feature. For a given feature point Fi, the relative angle to every other feature point Fj (j≠i) is defined as ∠Fi O Fj, where O is the origin, or region centre. If a boundary is rotated or scaled the features and relative angles remain constant. Thus this representation exhibits properties of rotation and scale invariance.

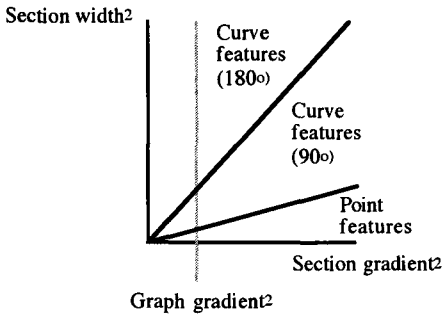


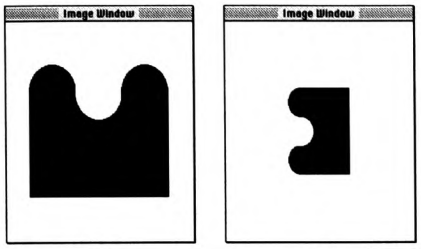
Figure 5. Partitions of feature space for classification of feature graph sections.

4 Experimental Results

The neural network pipeline has been implemented in C on the Apple Macintosh Performa 630 system. Images of both biological and simple geometric shapes have been tested. Figure 6(a) shows a simple geometric shape which is presented to the feature extraction network. The resulting feature graph is shown in figure 7(a). Figure 6(b) shows an image of the same shape but rotated 90 degrees counter clockwise and scaled by approximately 50%. The feature graph derived from the data in the neural network pipeline is shown in figure 7(b). The vertical lines show where the graph, or boundary is split into its component features. Since the latter shape has been rotated the feature graph in figure 7(b) is "out of phase" with the feature graph in figure 7(a). However, visual inspection reveals similarity between both feature graphs and a correspondence can be found between the features in each graph. This shows that the network can output data which exhibits properties of rotation and scale invariance for simple geometric shapes.

Figure 8(a) shows an image containing

two fungal spores, *Pestalotiopsis japonica* (top right) and *Truncatella truncata* (bottom left). Both specimens are extracted from the pipeline and their boundaries are shown in figures 8(b) and 8(c) respectively. The feature graphs for each of these specimens are shown in figures 9(a) and 9(b). Both graphs have been smoothed out with a noise reduction algorithm because the noise contained in the original boundaries produced a spiking effect in the graph.



(a) (b)

Figure 6. Geometric shapes

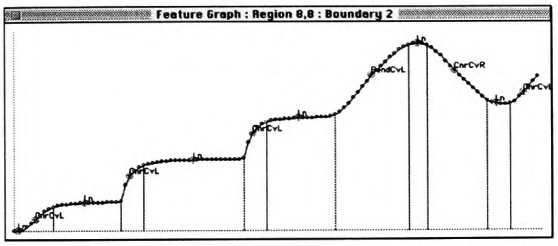


Figure 7(a). Feature graph for shape given in figure 6(a)

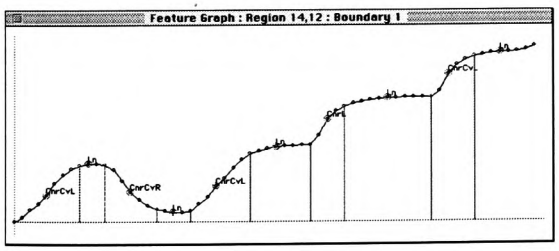
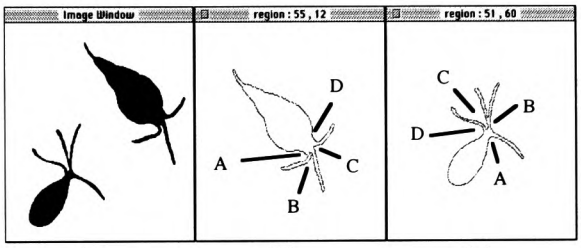


Figure 7(b). Feature graph for shape given in figure 6(b)

Unlike the feature graphs for the geometric shapes, finding corresponding features between these two graphs is not straightforward. Firstly, the configuration of the appendages on each specimen is different, that is, the layout and number of appendages varies between each specimen. Secondly, each specimen contains a great deal of variance in the direction change of their respective boundaries which results in over-segmentation of the feature graph, as shown in figures 9(a) and 9(b). Finally, because the appendages vary in position, size and number, finding a common starting point on the boundary can be difficult. This results in the feature graphs in figures 9(a) and 9(b) also being "out of phase". We can begin to look for correspondence by analysing the global features of each graph. The most significant features are the troughs caused by the concavity of the boundary where the appendages meet the main body of the specimen. These are labelled in figures 8(b), 8(c), 9(a) and 9(b) so correspondences between different sections of the feature graphs can be seen. However, because of over-segmentation, the global properties of each boundary section may not be described which makes finding corresponding features between each graph more difficult.



(a) (b) (c)

Figure 8. Images of *P.japonica* and *T.truncata*

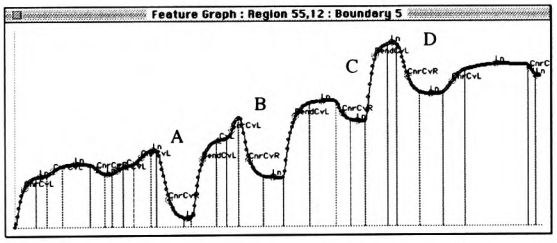


Figure 9(a). Feature graph for *P.japonica*

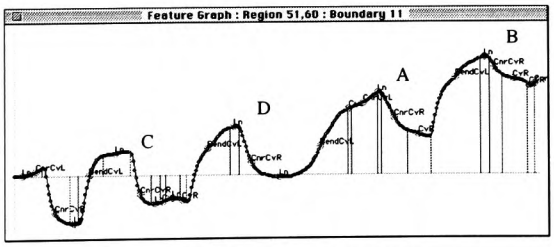


Figure 9(b). Feature graph for *T.truncata*

5 Conclusion & Future Work

The neural network pipeline proposed here has been implemented and tested using images of geometric and biological shapes. The network can successfully extract high level features from simple grey-scale and

binary images. These features have been used to form a region hierarchy where the representation of each boundary adopts a description-by-parts approach [2]. This works well for geometric forms where each boundary can be described in terms of lines, curves, corners and points. For biological images which contain irregular shapes, over-segmentation of the resulting feature graph makes finding corresponding features difficult. Morphologically, one can describe an object with landmark points at significant locations on the boundary, or by describing the boundary by fitting a function to the outline [10]. The biological specimens tested provide few significant landmarks, so future work will look at modelling the outline with parametric cubic spline functions, firstly because they can be used to describe all or part of the boundary and secondly because they overcome the limitation of the basic cubic spline function [10]. Finally, it is desirable that the network process more complex images. To this end, it is proposed the pipeline structure be changed so the RL layer can first be used to segment the image, using a more complex homogeneity function to overcome noise and shading, from which the H and V layers can extract the edge map for further processing.

References

[1] Adams, R. & Bischof, L. Seeded region growing. IEEE Trans. on Pattern Analysis and Machine Intelligence. 16(6): 641--647, 1994.

[2] Ben-Arie, J., Rao, K. R. & Wang, Z. A neural network approach for shape description and invariant recognition. ARPA Image Understanding Workshop (IUW) Conference Proceedings: 863--870, 1994.

[3] Bhandarkar, S.M., Koh, J. & Suk, M. A hierarchical neural network and its application to image segmentation. Mathematics and Computers in Simulation. 41: 337--355, 1996.

[4] Fukushima, K. Analysis of the process of visual recognition by the Neocognitron. Neural Networks. 2: 413--420, 1989.

[5] Horn, B.K.P. Robot Vision. MIT Press, 1986

[6] Kohonen, T. Self-organisation and associative memory, 2nd Edition. Springer, Berlin, 1988.

[7] Kulkarni, A.D. Artificial neural networks for image understanding. Van Nostrand Reinhold NY, 1994.

[8] Lin, W-G. & Wang, S-S. A new neural model for invariant pattern recognition. Neural Networks. 9(5): 899--913, 1996.

[9] Low, A. Introductory computer vision and image processing. McGraw-Hill London, 1991.

[10] Rohlf F.J. & Bookstein F.L. (eds). Proceedings of the Michigan morphometrics workshop, 1990.

[11] Srinivasa, N. & Jouaneh, M. A neural network for invariant pattern recognition. IEEE Trans. on Signal Processing. 40: 1595--1599, 1992.

[12] Srinivasa, N. & Jouaneh, M. An invariant pattern recognition machine using a modified ART architecture. IEEE Trans. on System, Man and Cybernetics. 5: 1432--1437, 1993.

[13] Widrow, B. & Winter, R. Neural nets for adaptive filtering and adaptive pattern recognition. Computer. 21: 25--39, 1988.

[14] Widrow, B. *et al*. Layered neural nets for pattern recognition. IEEE Trans. on Acoustics Speech and Signal Processing. 26: 1109--1118, 1988.

[15] Zeki, S. The visual image in mind and brain. Scientific American. 267(3): 43--50, 1992.

[16] Zheng, Y-J. Feature extraction and image segmentation using self-organising networks. Machine Vision and Applications. 8: 262--274, 1995.

Analysing the Mallat Wavelet Transform to Delineate Contour and Textural Features

Mr. Paul Angel
School of Computing, The University of Glamorgan

and

Mr. Colin Morris
School of Computing, The University of Glamorgan

Wavelets to Delineate Edge and Texture Features

Contact:

Paul Angel
School of Computing
The University of Glamorgan
Pontypridd, South Wales, UK. CF37 1DL.

Tel: +44 (0)1443 482731
Fax: +44 (0)1443 482715

email: pangel@glam.ac.uk

Abstract

Analysing natural scenes is made difficult when both contour and textural features are present. The problem is compounded when analysing images of biological specimens where natural deformation means landmark features will not always correspond from one image to the next. Separating contour and textural features presents a non-trivial problem in that information is needed about which parts of the image contain texture. The Mallat wavelet transform is analysed to separate contour and textural features. The emergence of events in a coarse to fine scale traversal, the wavelet transform modulus maxima and the local activity surrounding each wavelet response are deemed important characteristics in distinguishing contour and textural features. An algorithm is described which uses these features to extract edges at each scale and combine them to form a single contour map. This is then applied to synthetic and electron-microscopic images of biological specimens containing texture and contour, and contour-only features. Initial results are promising, showing that contour and textural features can be separated.

List of mathematical symbols used throughout the paper

Symbol	Name / Font used
+	Addition (Font : Times ; Size 12pt.)
*	Multiplication / Convolution (Font : Times ; Size 12pt.)
δ	Delta (lower case) (Font : Symbol ; Size 10pt.)
∂	Delta (partial derivative) (Font : Symbol ; Size 10pt.)
ε	Epsilon (lower case) (Font : Symbol ; Size 10pt.)
=	Equals (Font : Times ; Size 12pt.)
$\neq$	Not equal to (Font : Times ; Size 12pt.)
∞	Infinity (Font : Symbol ; Size 10pt.)
λ	Lambda (lower case) (Font : Symbol ; Size 10pt.)
ψ	Psi (lower case) (Font : Symbol ; Size 10pt.)
π	Pi (lower case) (Font : Symbol ; Size 10pt.)
$\int$	Integration (Font : Symbol ; Size 10pt.)
σ	Sigma (lower case) (Font : Symbol ; Size 10pt.)
-	Subtraction (Font : Times ; Size 12pt.)
τ	Tau (lower case) (Font : Symbol ; Size 10pt.)
Θ	Theta (upper case) (Font : Symbol ; Size 10pt.)
θ	Theta (lower case) (Font : Symbol ; Size 10pt.)

1. Introduction

Edge detection and texture analysis are two fundamental low-level image feature extraction processes, suited to synthetic and natural scene analysis respectively [23]. Contour features contain invaluable shape information about objects in a scene. Attneave observed that most of the information about object shape is contained at dominant points of high curvature [1]. Texture is an equally important feature, describing the internal structure, or surface of objects [36]. Generally, image feature extraction techniques focus on either edge detection or texture analysis, but not both [23]. Ganesan and Bhattacharyya [9] have addressed the problem of detecting edges in images containing contour or textural features, but they do not address the problem of distinguishing between these types of features. To do so requires information about which parts of the image contain texture.

Scale space provides a possible solution in that an image can be analysed according to a specific context, or scale, allowing features of sufficiently small spatial extent to be separated from larger scale features. Multiscale image analysis delineates features according to their spatial scale by smoothing, or blurring out smaller scale features [4]. Nonlinear approaches vary the blurring so important structures such as edges are not eliminated [24]. Wavelet analysis provides a powerful set of tools for building multiscale representations [5] and have received much attention throughout the machine vision literature [6,7,17,25], first because they have compact support allowing local image features to be analysed [4], and second because neurophysiological evidence indicates that receptive fields in the primary visual cortex and peripheral regions such as the retina approximate wavelet functions and perform similar filter operations [11,21].

Wavelets have been applied to both edge detection and texture analysis. Edges are extracted at different scales by analysing the image with dilations of a "mother wavelet", usually the first or second order derivative of the Gaussian smoothing function given the stability of the resulting multiscale representation [15,34]. The Marr wavelet which approximates the second derivative of the Gaussian and commonly referred to as the Marr/Hildreth operator, has been used extensively for edge detection [22]. The Mallat wavelet which approximates the first derivative

of the Gaussian has also been used for edge detection [18,19], the response of which is equivalent to Canny's algorithm [3], and hence optimal in terms of Canny's detection, localisation and one response criteria [19]. Although in terms of Canny's criteria, optimal second derivative operators exist [30], it is the actual properties of the zero crossings in the response which make first derivative operators more attractive. In the modulus of the first derivative response, maxima correspond to sharp inflection points in the Gaussian smoothed image while minima correspond to weak inflection points. Zero crossings correspond to both, making it difficult to distinguish between salient and non-salient features in the image. Also, the size of the maxima can be used to differentiate between small and large amplitude fluctuations. Zero crossings only retain positional information and as such, these features cannot easily be distinguished [19]. Texture analysis requires a different set of wavelet tools which analyse the local orientation and spatial frequency of the image within a conjoint spatial-spectral representation [6,7,25,26,27,33]. Textural properties can be analysed from the response of a wavelet filter bank configured to cover an adequate orientation and spatial-frequency range [6]. The Gabor wavelet, a special case of the Morlet wavelet [36], has been widely used in texture classification and segmentation problems, while the application of critically sampled and adaptive filters improves the computational requirements and classification accuracy [25]. No one wavelet has been used for both edge detection and texture analysis, although the Gabor can be configured as an odd, asymmetrical function which approximates the first derivative of the Gaussian [23] and the statistical distributions of edge responses can been used to classify textural regions [31].

In analysing natural scenes where both contour and textural features exist, texture analysis wavelets will extract the local orientation and spatial frequency properties of the image while edge detection wavelets will extract information about the contours in the image. The same grey level variance that defines each edge also defines the textural features, or micro edges within the image [23,31]. This introduces the problem of how to differentiate between these features, to determine if a local wavelet response represents the presence of a contour or the properties of local textural features. Understanding the nature of texture presents a non-trivial problem and no attempt is made here to offer a formal, unified definition of texture. The properties of the multiscale representation resulting from the Mallat wavelet transform are

studied and applied to the analysis of electron-microscopic (EM) images of biological specimens, a special case of natural scene analysis. It was noted above that both contour and textural features are important in describing objects in a scene, the former more so, especially in the field of morphometrics, where key landmark features used for species classification reside on object contours [2,13]. In analysing electron-microscopic images such as that given in figure 3, the Mallat wavelet is used to extract contour features across a suitable scale range. Since the grey level variance in the image can correspond to both edges and texture, the Mallat wavelet interprets textural features as a collection of edges. It is important that these are removed so salient contour features can be identified. The emergence of edge events in a coarse to fine scale traversal along with the properties of the wavelet transform modulus maxima and the local activity along the gradient of each wavelet response are deemed important characteristics in distinguishing contour and textural features.

The paper is organised as follows. Section 2 describes the Mallat wavelet transform and the properties of the resulting multiscale representation. The problem of localising displaced contours at larger scales and the elimination of false edges is then discussed. The algorithm to separate contour and textural features is then described in section 3. The results of the algorithm applied to synthetic and electron-microscopic images containing both contour and textural, and contour-only features are discussed in section 4 along with issues relating to the scalability of the algorithm, while the conclusions are discussed in section 5.

2. Multiscale Image Analysis

2.1. Mallat Wavelet Decomposition

A wavelet can be any function $\psi(x)$ that satisfies the condition $\int_{x=-\infty..\infty} \psi(x)\, dx=0$. The Mallat wavelet is a quadratic spline approximation of the first derivative of the Gaussian smoothing function $\theta(x)$ which is sufficiently localised in time and frequency [4,19].

$$\psi(x) = \frac{d\theta(x)}{dx} \qquad (1)$$

For one dimensional signals, the wavelet transform at scale s, $W_s f(x)$, is obtained by convolving dilations of the "mother wavelet" $\psi_s(x)$ where

$$\psi_s(x) = \frac{1}{s}\,\psi\left(\frac{x}{s}\right) \tag{2}$$

with the signal f(x).

$$W_s f(x) = f * \psi_s(x) \tag{3}$$

The one dimensional wavelet transform can be extended into two dimensions using two orthogonal wavelets $\psi^1(x,y)$ and $\psi^2(x,y)$ approximating the first derivative of the two dimensional Gaussian smoothing function $\theta(x,y)$.

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \qquad \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y} \tag{4}$$

For two dimensional images, the wavelet transform at scale s contains two components, W_s^1 and W_s^2, obtained by convolving the wavelet dilations $\psi_s^1(x)$ and $\psi_s^2(x)$ where

$$\psi_s^1(x,y) = \frac{1}{s^2}\,\psi^1\left(\frac{x}{s},\frac{y}{s}\right) \qquad \psi_s^2(x,y) = \frac{1}{s^2}\,\psi^2\left(\frac{x}{s},\frac{y}{s}\right) \tag{5}$$

with the image f(x,y).

$$W_s^1 f(x,y) = f * \psi_s^1(x,y) \qquad W_s^2 f(x,y) = f * \psi_s^2(x,y) \tag{6}$$

The gradient vector $M_s f(x,y)$ indicating the direction of the combined wavelet response in the image plane is defined as

$$M_s f(x,y) = \sqrt{|W_s^1 f(x,y)|^2 + |W_s^2 f(x,y)|^2} \tag{7}$$

and the orientation of the gradient vector $A_s f(x,y)$ is given by

$$A_S f(x,y) = \tan^{-1}\left(\frac{W_S^2 f(x,y)}{W_S^1 f(x,y)}\right) \quad \text{where } 0 \leq A_s f(x,y) < 2\pi. \tag{8}$$

For one dimensional signals, maxima in the modulus of the wavelet response correspond to sharp gradient changes. In two dimensions, local maxima along the gradient of the combined wavelet response correspond to individual edge points [19]. This is equivalent to the output of Canny's algorithm and hence optimal in terms of Canny's detection, localisation and one response criteria [3]. The wavelet transform modulus maxima contains important information about contours in the image. Mallat and Zhong show that maxima decay across multiple scales characterises different types of edge [19]. The behaviour of maxima across each scale is important, and to investigate this, a set of isolated one dimensional step edge functions $F_{\sigma A}(x)$ are analysed where

$$F_{\sigma A}(x) = \begin{cases} \begin{pmatrix} -A \text{ if } x < 0 \\ A \text{ if } x \geq 0 \end{pmatrix} & \text{if } \sigma = 0 \\ \dfrac{A}{1+e^{\frac{-x}{\sigma}}} & \text{otherwise} \end{cases} \tag{9}$$

and σ and A represent respectively the continuity and amplitude of the step edge. The wavelet transform as defined in (3) is equivalent to the cross-correlation of the wavelet dilation at scale s, $\psi_s(x)$, with the step edge function $F_{\sigma A}(x)$. The wavelet response at each scale indicates the relevance of each event at that scale, and only begins to decay when the gradient of the wavelet $\delta\psi_s(x)$ is greater than the gradient of the step edge $\delta F_{\sigma A}(x)$. This results in the characteristic positive decay where the maxima values increase proportionately to the scale [19]. Negative decay results when isolated events with positive or zero decay are analysed in the global context of the signal. For each scale of analysis, s, negative decay results when

$$\frac{R_{s+1}(x)}{M_{s+1}} < \frac{R_s(x)}{M_s} \tag{10}$$

where $R_s(x)$ is the wavelet response for event x at scale s and M_s is the maximum wavelet response at scale s. The spatial extent of each singularity defined by the events in the signal is also important in determining the decay of maxima across multiple scales. Figure 1(a) shows a one dimensional signal extracted from the image in figure 3, with the spatial extent of the singularity at abscissa 2 increasing monotonically. The wavelet transform modulus maxima exhibit negative decay for the events that define this singularity, which inverts to positive decay as the spatial extent increases, as shown in figure 1(b). Given the cross correlation properties of the wavelet transform, the maxima response indicates the relevance of events at each scale, and given that maxima exhibit negative decay only in the global context of each signal, the maxima at the largest scale of support, or the scale at which the event emerges in a coarse to fine scale traversal, is most relevant.

In creating a multiscale representation, it is important that sufficient information about the original image is retained, although for image feature extraction this is less critical than for signal reconstruction [16]. Hence the choice of scale range is important. It is well known in information theory that logarithmic sampling leads to informationally uniform sampling density [8]. Therefore a dyadic scale sequence is used to retain most of the information in a signal, even though it is over-complete, retaining redundant information [19]. Discrete images have a limited spatial scale range, bounded by the sampling resolution and the spatial extent of the features within the image [14]. For electron-microscopic images such as that given in figure 3, salient landmark information is contained within the dyadic scale range $S_2{}^0$ to $S_2{}^3$, and throughout this paper, the Mallat wavelet transform and subsequent analysis is confined to this range.

2.2. Localising Displaced Maxima

In order to determine the characteristics of wavelet responses across multiple scales it is important to find correspondences between maxima at each scale of analysis. In creating a multiscale representation, maxima at larger scales do not occur at their true location [8,34]. This displacement results from the blurring required to eliminate smaller scale features which is

determined by the width of the underlying Gaussian smoothing function, and the interaction with surrounding events [32]. For one dimensional signals, individual points are displaced creating characteristic scale space arches [34]. In two dimensions, individual points become contours which deform, split and merge as the scale becomes larger [8,15,16]. The problem becomes more tractable when an image is analysed across a sufficiently dense range of scales so the displacement of edge points is bounded by 1 pixel [16,19]. However, this is computationally expensive.

Numerous solutions have been considered throughout the literature which localise maxima or zero crossings. For one dimensional signals, Mallat and Zhong propose an algorithm where maxima of sufficient amplitude are connected to the closest maxima at the next largest scale which has the same sign in the wavelet transform [19]. Ulupinar and Medioni analyse the displacement of zero crossings which correspond to isolated step edge events, looking for correspondences across multiple scales as well as analysing LoG responses at individual scales to determine their true location [32]. Gokmen and Jain propose a general edge detector scheme where filters become a special case in $\lambda\tau$-space. λ represents the scale while τ represents the continuity of the filter around x=0. As τ tends to zero, the filter becomes more discontinuous and the displacement of maxima at larger scales is reduced, but sensitivity to noise is increased. As τ tends to 1, filters become more continuous, but the location of maxima at larger scales is sacrificed for more robust handling of noise. Gokmen and Jain recommend a compromise where $\tau \approx 0.5$, reducing the displacement of filter responses while retaining a moderate handling of noise [10]. In two dimensions, correspondences are found between contours rather than individual edge points [17,32]. Niessen *et al* consider linking events in a hyperstack scale space representation based on their intensity [24]. Lu and Jain's RESS system [16] employs the CEL (Correcting Edge Location) procedure which finds correspondences between individual edge points, not contours, which is made more difficult given the splitting, merging and deformation that results at larger scales [16,8]. Also, in considering the presence of textured regions, it is not meaningful to consider textural features as contours. Lu and Jain's CEL procedure looks for corresponding zero crossings at adjacent scale levels within a circular region of interest bounded by the Gaussian width σ. Correspondences are found between zero crossings and their closest neighbours at adjacent

scales, that is those zero-crossings with the smallest radial distance. Here a similar approach is adopted to find correspondences between maxima in the Mallat wavelet transform. However, neighbouring edge points on the same contour should be ignored and only maxima along the direction of the wavelet response are considered. Correspondences are found according to the relative radial and angular positions as well as the intensity of the maxima. Due to the splitting and merging of contour points that occurs in two dimensional scale space and the emergence of false edges at larger scales, there is no guarantee that a 1:1 correspondence exists between maxima at adjacent scales. As a result discontinuities occur in the localised contours at larger scales. Also, given the sparse analysis density around larger scales in the dyadic range described above, errors in the correspondences found are likely to occur since information from intermediate scales of analysis is missing. These problems can be overcome by iteratively creating and evaluating correspondences found between maxima. Where more than one possible correspondence exists between two scales, a selection is made as to the best correspondence based on the above position and intensity criteria. The problem of discontinuous contours at larger scales can be resolved by looking for support at smaller scales.

2.3. Identifying and Eliminating False Contours

A major problem which occurs with the splitting and merging of contour points is that false edges can emerge at larger scales of analysis [8,16,35]. This is an unavoidable effect of Gaussian based edge detection operators. One of the important properties of Gaussian derived multiscale representations is that all contours can be tracked to their true location at the smallest scale of analysis [34]. False edges can thus be identified when they have no support at smaller scales. Thus, once maxima have been localised, false edge points can be eliminated by removing points which have no corresponding maxima at the adjacent smaller scale. Applying this rule in a fine to coarse scale traversal allows false edges that emerge at any given scale and perpetuate at larger scales to be eliminated.

3. Delineating Contour and Textural Features

Understanding the nature of texture presents a non-trivial problem in that it is difficult to define what actually constitutes textural features. Though no formal definition of texture is attempted, the features deemed important in discriminating between contour and textural features are the context in which the image is analysed, the wavelet transform modulus maxima and the local activity along the gradient of the wavelet response. Context is important, and in the study of contour and textural features, it is considered a duality exists in that features can be perceived as either. Multiscale representations, such as that created with the Mallat wavelet transform described above, allow images to be analysed within different contexts. Consider the "brick wall" image in figure 2(a) and the resulting Mallat wavelet transform in figure 2(b). At smaller scales of analysis, the features defining each brick are present while in the context of the largest scale, the brick features are eliminated and the wall as a whole can be seen. Therefore, as the context changes and the scale of analysis becomes larger, so texture emerges. Hence, texture does not exist at the scale its constituent features reside, but only emerges in the context of larger scales of analysis. The image in figure 2(a) presents an idealised case where all of the features below the largest scale of analysis can be considered textural features.

Natural scenes can contain important contour information at any scale, so considering only the largest scale of analysis may ignore important features contained in the multiscale representation. To overcome this problem, an algorithm is described which analyses the emergence of events in a coarse to fine scale traversal, and combines the multiscale representation created with the Mallat wavelet transform to produce a single contour map where textural features are filtered out. This uses the wavelet transform modulus maxima at the largest scale of support along with the local activity surrounding each event. This is determined by both the spatial frequency of the texture and the context in which it is analysed. Homologous regions are defined by isolated contour events with little or no surrounding activity. Depending on the context of analysis, textured regions can be bounded by texture of different spatial frequency and orientation [6] or by smooth regions. The latter results in the decay of local activity around contour points while the former presents a class of boundary which cannot easily be identified from the wavelet transform itself, and as such requires a two-

phase approach where texture region segmentation is applied to the textured regions identified in the contour-texture delineation phase described below.

For each maxima $M_2{}^j{}_{xy}$ that emerges at scale $S_2{}^j$, the local activity, or interaction with neighbouring events, $I_2{}^j{}_{xy}$, is calculated based on the relative position and intensity of neighbouring maxima within the square region of interest $\varepsilon x \varepsilon$, centred on $M_2{}^j{}_{xy}$, where ε is defined as

$$\varepsilon = \frac{1}{4} 2^{(2^{n-1} - 2^j)} \tag{11}$$

and n equals the number of scales of analysis in the initial wavelet transform. Since maxima can correspond to individual edge points, neighbouring maxima on the same contour are not relevant so only those maxima which lie along the axis specified by the orientation of the gradient of the wavelet response are considered. This axis is defined by $A_2{}^{j+1}{}_{xy}$, given in (8) and $A'_2{}^{j+1}{}_{xy}$ which is defined as

$$A'_2{}^{j+1}{}_{xy} = \begin{cases} A_2{}^{j+1}{}_{xy} + \pi & \text{if } A_2{}^{j+1}{}_{xy} < \pi \\ A_2{}^{j+1}{}_{xy} - \pi & \text{otherwise.} \end{cases} \tag{12}$$

For each neighbouring maxima $M_2{}^j{}_{x'y'}$, where (x',y') belongs to the set of points

$$\{x',y' : x-(\varepsilon/2) \leq x' < x+(\varepsilon/2) \; ; \; y-(\varepsilon/2) \leq y' < y+(\varepsilon/2) \} \tag{13}$$

the radial distance $\lambda_{x'y'}$ from $M_2{}^j{}_{xy}$ is defined as

$$\lambda_{x'y'} = \sqrt{|x-x'|^2 + |y-y'|^2} \tag{14}$$

while the orientation of $M_2{}^j{}_{x'y'}$ w.r.t. $M_2{}^j{}_{xy}$ is defined as

$$\theta_{x'y'} = \tan^{-1}\left(\frac{y-y'}{x-x'}\right) \qquad \text{where } 0 \le \theta_{x'y'} < 2\pi. \tag{15}$$

Let Θ be the set of all relevant points $M_2{}^j{}_{x'y'}$ that satisfy the inequality

$$\delta\theta_{x'y'} < \text{MAX}\left(\frac{2\pi}{8\,(\text{MAX}(|x-x'|,|y-y'|))}, t\right) \tag{16}$$

where t is a suitable orientation threshold defining the angular window of interest perpendicular to the axis given in (8) and (12), and $\delta\theta_{x'y'}$ is the orientation offset of neighbouring maximum $M_2{}^j{}_{x'y'}$ relative to this axis which is defined as

$$\begin{aligned}\delta\theta_{x'y'} = \text{MIN}\,(\; & |\,\theta_{x'y'} - A_2{}^{j+1}{}_{xy}\,|\,,\ |\,\theta_{x'y'} - (A_2{}^{j+1}{}_{xy} + 2\pi)\,|\,,\\ & |\,\theta_{x'y'} - (A_2{}^{j+1}{}_{xy} - 2\pi)\,|\,,\ |\,\theta_{x'y'} - A'_2{}^{j+1}{}_{xy}\,|\,,\\ & |\,\theta_{x'y'} - (A'_2{}^{j+1}{}_{xy} + 2\pi)\,|\,,\ |\,\theta_{x'y'} - (A'_2{}^{j+1}{}_{xy} - 2\pi)\,|\;)\end{aligned} \tag{17}$$

where MAX and MIN are functions which return the argument with the largest and smallest value respectively. Given the orientation inaccuracy around neighbouring pixels inherent in square tessellated discrete images, if $t < \pi/4$ then important neighbouring pixels can be ignored. Hence the argument $2\pi/8(\text{MAX}(|x-x'|,|y-y'|))$ in (16) ensures neighbouring pixels in close proximity to $M_2{}^j{}_{xy}$ whose minimum angular distance is greater than t are considered. The interaction $I_2{}^j{}_{xy}$ of $M_2{}^j{}_{xy}$ with the relevant neighbouring maxima is calculated with a modified convolution operation where $M_2{}^j{}_{xy}$ is ignored.

$$I_2{}^j{}_{xy} = \sum_{\substack{y'=y-\varepsilon/2 \\ (y' \neq x)}}^{y' \le y+\varepsilon/2} \sum_{\substack{x'=x-\varepsilon/2 \\ (x' \neq x)}}^{x' \le x+\varepsilon/2} e^{\frac{-\lambda_{x'y'}{}^2}{4\varepsilon}}\, M_2{}^j{}_{x'y'} \qquad \text{where } (x',y') \in \Theta \tag{18}$$

The contour response $C_2{}^j{}_{xy}$ for each event $M_2{}^j{}_{xy}$ that emerges at scale $S_2{}^j$ combines the

interaction of the event with the wavelet transform modulus maximum and is defined as

$$C_2{}^j{}_{xy} = \frac{1}{e^{\left(\frac{I_2{}^j{}_{xy}}{MI_2{}^j}\right) c}} M_2{}^j{}_{xy}$$

$$\text{where } 0 \leq C_2{}^j{}_{xy} \leq 1 \qquad (19)$$

where $MI_2{}^j$ is the largest interaction value for maxima that emerge at scale $S_2{}^j$ and c is a scaling factor which determines the relevance of the interaction with neighbouring events. When $c \equiv 4$, isolated events produce a large contour response while non-isolated events produce a small contour response. The contour responses at each scale are then thresholded to eliminate non-salient features and retain only those contour events that emerge at the respective scale. Experimentation has found that most information resides below a contour response of 0.125, while salient contour information resides above this value. Eliminating all contour responses below the threshold of 0.125 retains important contour features. The thresholded contour responses at each scale are then logically ORed to produce the final contour map.

4. Experimental Results

The algorithm described above has been applied to electron-microscopic images of biological specimens and synthetic images to test its effectiveness in retaining salient contour features while filtering out textural information. For each image, the Mallat wavelet transform is calculated, the displacement of maxima at larger scales is resolved and false edges which emerge are removed. For each scale of analysis, the contour response is calculated where c =4 and then thresholded using a threshold value of 0.125. The results are then combined to form a complete contour map of the image.

The first experiment analyses an electron-microscopic image of the head of an adult diptera (fly) shown in figure 3. This contains important contour information across all scales of analysis as well as textured regions. The localised wavelet transform maxima are shown in figure 4 while the contour responses for each of the localised maxima images are shown in figure 5. Combining each contour response produces the final contour map shown in figure 6.

The results show that the algorithm effectively eliminates regions of texture while retaining important contour events, including those that reside at smaller scales of analysis. Discontinuities or weak contour responses result for edge events that define features of small spatial extent such as the elongated structures present in figure 3. Although no activity exists around these features, two edge events in close proximity results in high interaction, making it difficult to distinguish between maxima of large intensity which are in close proximity and a textural region of weak intensity. The second electron-microscopic image analysed is that of the head of a Bull Ant, shown in figure 7. This contains primarily contour features, although non-salient detail exists within the image. The thresholded and combined contour map is shown in figure 8. The results shows that important contour features can be retained when no texture is present, and non-salient features within the image can be eliminated. The final image analysed is a synthetic image shown in figure 9(a). This contains both contour and textural features of varying spatial frequency. The resulting contour map is shown in figure 9(b). The results for this image are promising in that the textured regions of different spatial frequency are eliminated while retaining the important contour features. Discontinuities arise within certain edges as a result of the interaction between maxima in close proximity. Features on the perimeter of the textured region are also identified as contours. This results from the decay in interaction around the around the border of textured regions which neighbour continuous regions.

The Gabor approach has also been considered along with nonlinear multiscale analysis. The Gabor wavelet has been used successfully throughout the literature for texture region segmentation [6,26,27]. To determine if contour and textural features can be delineated from the Gabor wavelet response, a simplified bank of Gabor wavelets consisting of two orientations at 0 and 90 degrees, and two spatial frequencies has been applied to the electron-microscopic image in figure 3, the results of which are shown in figure 10. Unfortunately, the Gabor produces a non-localised response for isolated contour features, resulting in multiple maxima for a single edge. Since textural regions also produce a non-isolated response, the Gabor cannot be used to differentiate between contour and textural features.

The Mallat wavelet transform given in (6) constructs a linear multiscale representation where

the underlying Gaussian smoothing function at each scale of analysis remains constant across the whole image. Nonlinear approaches vary the smoothing at each scale, bounded by the width of the Gaussian used for linear decomposition, to preserve important edge features while smoothing homologous regions [24]. Approaches to configure the width of the smoothing function include geometrical and gradient-based methods. Here the Mallat wavelet is locally configured according to the local activity surrounding each event, so textured regions will be blurred out while retaining contour features. First the gradient of the image at each point is calculated along with its orientation. Interaction is calculated according to (18), and each wavelet is configured accordingly, the dilation of the wavelet being proportional to the interaction. The results of the wavelet transform should then filter out textured regions. However, the results show that contour and textural features are not easily separated using this approach, as shown in figure 11. This approach is based primarily on the interaction within the gradient image. An important property of linear Mallat wavelet decomposition is that structures of sufficiently small spatial extent that satisfy the condition in (10) produce negative maxima decay in the resulting multiscale representation, which reduces the corresponding contour response. This information is not available in the non-linear approach described above. The maxima decay for the image in figure 3 is shown in figure 12. Here, textured regions exhibit negative decay, as do the contour events defining structures of small spatial extent, although they interact less allowing them to be identified as contour features, as shown in figure 6. The spatial extent of the textural features in the above images are confined to limited spatial frequencies, and provided the wavelet transform covers an adequate scale range which corresponds with the spatial extent of features within the image, the technique described above can be extended to images containing texture of greater spatial extent.

5. Conclusion

The problem of analysing images which contain both contour and textural features has been discussed. Since wavelet responses can correspond to both textural and contour features it is important that they are interpreted correctly. The Mallat wavelet response has been studied and the context in which features are interpreted, the local activity around each wavelet response and the wavelet transform modulus maxima are considered important features in distinguishing

contour and textural features. An algorithm has been developed into which these features are incorporated. This has been applied to synthetic images and electron-microscopic images of biological specimens which contain both contour and textural features. The algorithm has been compared against non-linear multiscale decomposition and Gabor analysis for delineating contour and textural features. The results of the algorithm developed above show that important contour features can be separated from textural regions. This performs better than non-linear analysis where textural features cannot easily be distinguished from contour features since information about the behaviour of events across multiple scales is not available, which is important in extracting contour features. Analysing images with a bank of Gabor wavelets produced poor results since both contour and textural features produce a non-localised response. Future work will look at overcoming problems inherent in the algorithm described above. In calculating the contour response at each scale, maxima in close proximity results in a large interaction coefficient, making them difficult to distinguish from textured regions of weak intensity. This problem could be overcome by analysing, for a given event, the increase in local activity across all scales. Also, a number of maxima within the border of textured regions can be identified as contour features. Since interaction decays around the perimeter of textured regions, true contour points could be found by looking for interaction minima. Finally, the spatial frequency and orientation information contained in the Mallat wavelet transform could be used to configure a bank of texture analysis filters which can be applied to the textured regions identified by the above algorithm.

References

1. F. Attneave, Symmetry, Information and Memory for Patterns, *American Journal of Psychology*. **68**, 1955, 209-222.

2. F.L. Bookstein, Shape and the Information in Medical Images : A Decade of Morphometric Synthesis, *Computer Vision and Image Understanding*. **66**, 1997, 97-118.

3. J. Canny, A Computational Approach to Edge Detection, *IEEE Trans. on Pattern Analysis and Machine Intelligence*. **8**, 1986, 679-714.

4. A. Cohen and J. Kovacevic, Wavelets : The Mathematical Background, *Proc. IEEE*. **84**, 1996, 514-522.

5. I. Daubechies, Where Do Wavelets Come From? - A Personal Point of View, *Proc. IEEE*. **84**, 1996, 510-513.

6. J. G. Daugman, Complete Discrete 2D Gabor Transforms by Neural Networks for Image Analysis and Compression, *IEEE Trans. on Acoustics, Speech and Signal Processing*. **36**, 1988, 1169-1179.

7. G. Earnshaw, *Investigation of Texture Based Segmentation using Gabor Filters,* Centre for Intelligent Systems, University of Plymouth, 1993.

8. M. Fidrich and J-P. Thirion, *Multiscale Extraction and Representation of Features from Medical Images*, INRIA - Sophia Antipolis, 1994.

9. I. Ganesan and P. Bhattacharyya, Edge Detection in Untextured and Textured Images - A Common Computational Framework, *IEEE Trans. Systems, Man and Cybernetics*. **27**, 1997, 823-834.

10. M. Gokmen and A.K. Jain, $\lambda\tau$-space Representation of Images and Generalised Edge Detector, *IEEE Trans. Pattern Analysis and Machine Intelligence.* **19**, 1997, 545-563.

11. I.E. Gordon, *Theories of Visual Perception 2nd Edition*, Wiley, Chichester, 1997.

12. D. Hubel and T. Wiesel, Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex, *J. Physiol. (London).* **160**, 1962, 106-154.

13. R. Kirsch, Computer Determination of the Constituent Structure of Biological Images, *Comput. Biomed.* **4**, 1971, 315-328.

14. J.J. Koenderink, The Structure of Images, *Biolog. Cybern.* **50**, 1984, 363-370.

15. T. Lindeberg, Scale-Space for Discrete Signals, *IEEE Trans. on Pattern Analysis and Machine Intelligence.* **12**, 1990, 234-254.

16. Y. Lu and R.C. Jain, Reasoning about Edges in Scale Space, *IEEE Trans. on Pattern Analysis and Machine Intelligence.* **14**, 1992, 450-468.

17. S. Mallat, Wavelets for a Vision, *Proc. IEEE.* **84**, 1996, 604-614.

18. S. Mallat and W.L. Hwang, Singularity Detection and Processing with Wavelets, *IEEE Trans. on Information Theory.* **38**, 1992, 617-643.

19. S. Mallat and S. Zhong, Characterisation of Signals from Multiscale Edges, *IEEE Trans. Pattern Analysis and Machine Intelligence.* **14**, 1992, 710-732.

20. B.B. Mandelbrot, *The Fractal Geometry of Nature,* W.H. Freeman, San Francisco, 1982.

21. D. Marr, *Vision,* W.H. Freeman, San Francisco, 1982.

22. D. Marr and E. Hildreth, Theory of Edge Detection, *Proc. Royal Soc (London).* **207**, 1980, 187-217.

23. R. Mehrotra, K.R. Namuduri and N. Ranganathan, Gabor Filter-Based Edge Detection, *Pattern Recognition.* **25**, 1992, 1479-1494.

24. W.J. Niessen *et al*, Nonlinear Multiscale Representations for Image Segmentation, *Computer Vision and Image Understanding.* **66**, 1997, 233-245.

25. O. Pichler *et al* , A Comparison of Texture Feature Extraction Using Adaptive Gabor Filtering, Pyramidal and Tree Structured Wavelet Transforms, *Pattern Recognition.* **29**, 1996, 733-742.

26. M. Porat and Y.Y. Zeevi, The Generalised Gabor Scheme of Image Representation in Biological and Machine Vision, *IEEE Trans. on Pattern Analysis and Machine Intelligence.* **10**, 1988, 452-467.

27. M. Porat and Y.Y. Zeevi, Localised Texture Processing in Vision : Analysis and Synthesis in Gaborian Space, *IEEE Trans. Biomed. Engineering.* **36**, 1989, 115-129.

28. G. Ravichandran and M. Trivedi, Circular-Mellin Features for Texture Segmentation, *IEEE Trans. on Image Processing.* **4**, 1995, 1629-1640.

29. J.C. Russ, *The Image Processing Handbook,* CRC Press, 1992.

30. S. Sarkar and K.L. Boyer, Optimal Infinite Response Zero Crossing Based Edge Detectors, *CVGIP:Image Understanding.* **54**, 1991, 224-243.

31. M. Sonka, H. Vaclav and R. Boyle, *Image Processing, Analysis and Machine Vision,* Chapman & Hall, London, 1993.

32. F. Ulupinar and G. Medioni, Finding Edges Detected by a LoG Operator, *Computer Vision, Graphics and Image Processing.* **51**, 1990, 275-298.

33. M. Unser, Texture Classification and Segmentation Using Wavelet Frames, *IEEE Trans. on Image Processing.* **14**, 1995, 1549-1560.

34. A.P. Witkin, Scale Space Filtering, in *Proc. Int. Conf. AI.* ; 1983, pp.1019-1022.

35. A.L. Yuille and T.A. Poggio, Scaling Theorems for Zero Crossings, *IEEE Trans. on Pattern Analysis and Machine Intelligence.* **8**, 1986, 15-25.

36. Y.M. Zhu and R. Goutte, Analysis and Comparison of Space/Spatial Frequency and Multiscale Methods for Texture Segmentation, *Optical Engineering.* **34**, 1995, 269-282.

Figure Legends

Figure 1. (a) One dimensional signal where the singularity at abscissa 2 increases monotonically. (b) Maxima decay of the events in the wavelet transform defining the singularity at abscissa 2 across the dyadic scale range $S_2{}^0$ to $S_2{}^3$. This inverts from negative decay at S_0 to positive decay as the spatial extent of the singularity increases.

Figure 2. (a) "Brick Wall" Texture. (b) Mallat wavelet transform of the brick texture across the dyadic scale range $S_2{}^0$ to $S_2{}^3$.

Figure 3. Electron-microscopic image of the head of an adult diptera (fly). Species : *Drosophila Melanogaster* (Fruitfly). Copyright Dennis Kunkel PhD. University of Hawaii.

Figure 4. Localised wavelet transform modulus maxima for the image in figure 3 at scales (a) $S_2{}^0$, (b) $S_2{}^1$, (c) $S_2{}^2$ and (d) $S_2{}^3$.

Figure 5. Contour responses for the image in figure 3 at scales (a) $S_2{}^0$, (b) $S_2{}^1$, (c) $S_2{}^2$ and (d) $S_2{}^3$.

Figure 6. Final contour map for the image in figure 3 after thresholding and merging the contour responses at each scale.

Figure 7. Electron-microscopic image of the anterior view of the head of a Bull Ant. Copyright Tony Romeo, Electron Microscope Unit, The University of Sydney.

Figure 8. Contour map for the Bull Ant image in figure 7 after thresholding and combining the contour responses at each scale of analysis.

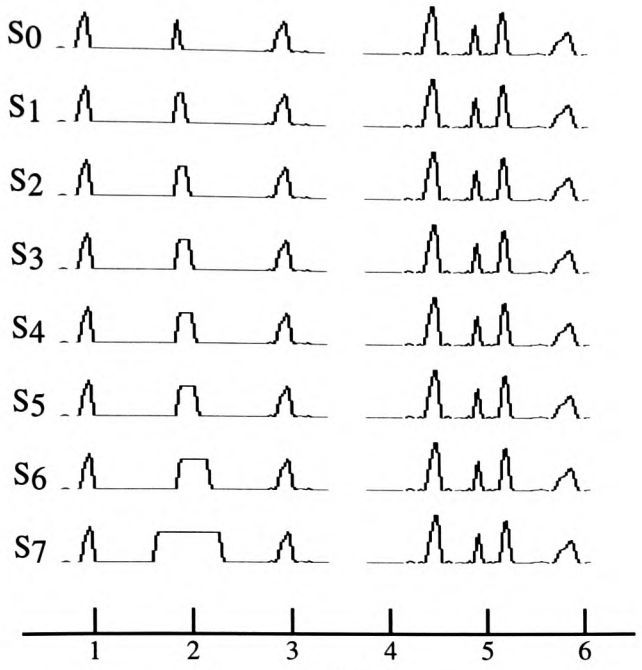
Figure 9. (a) Synthetic image containing contour and textural features of varying spatial

frequency. (b) Resulting contour map with the textural features filtered out while retaining important contour information.

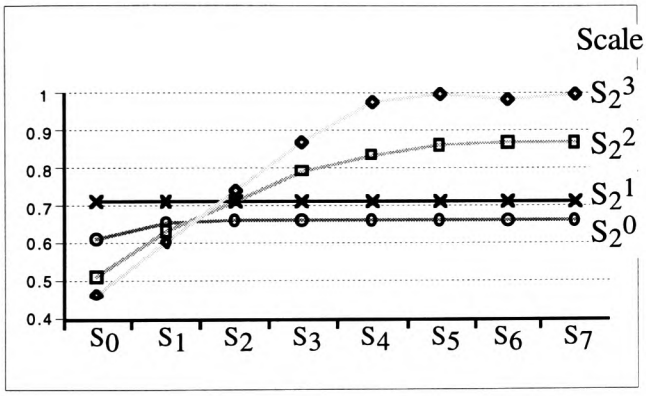
Figure 10. Gabor response for the image in figure 3, created with filters configured at 0 and 90 degrees and at scales S_2^2 and S_2^3.

Figure 11. Contour response produced by configuring the Mallat wavelet according to the local activity in the image given in figure 3. Textured regions cannot easily be separated from the contour features.

Figure 12. Decay of the wavelet transform modulus maxima from the image in figure 3. Light regions indicate negative decay, dark regions indicate positive decay while grey regions indicate zero decay. Textured regions and features of small spatial extent exhibit negative maxima decay.

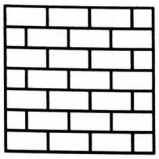

S0
S1
S2
S3
S4
S5
S6
S7
1
2
3
4
5
6

(a)

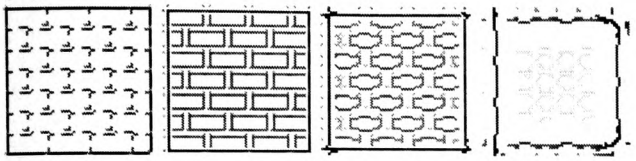

Scale
1
0.9
0.8
0.7
0.6
0.5
0.4
S0
S1
S2
S3
S4
S5
S6
S7
S2^3
S2^2
S2^1
S2^0

(b)

Figure 1



(a)



(b)

Figure 2

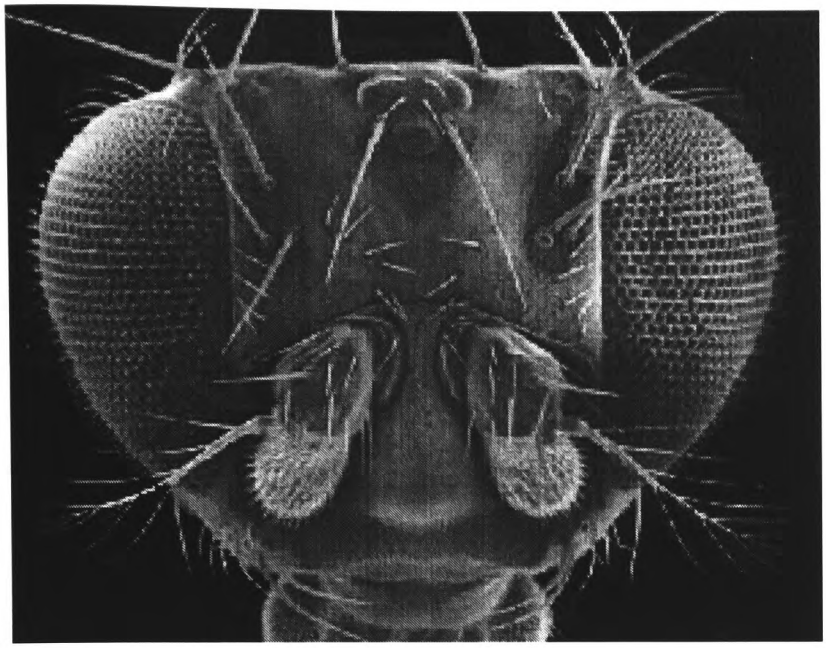


Figure 3

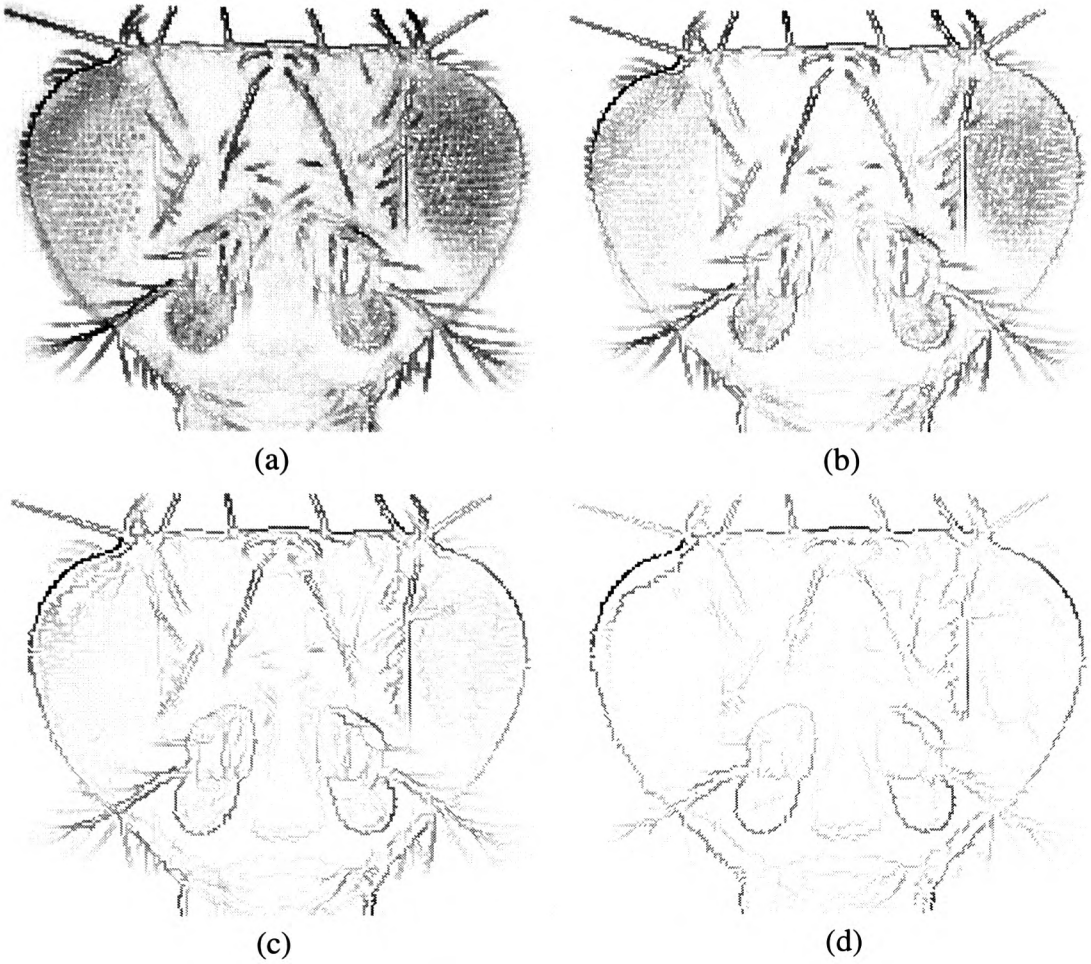
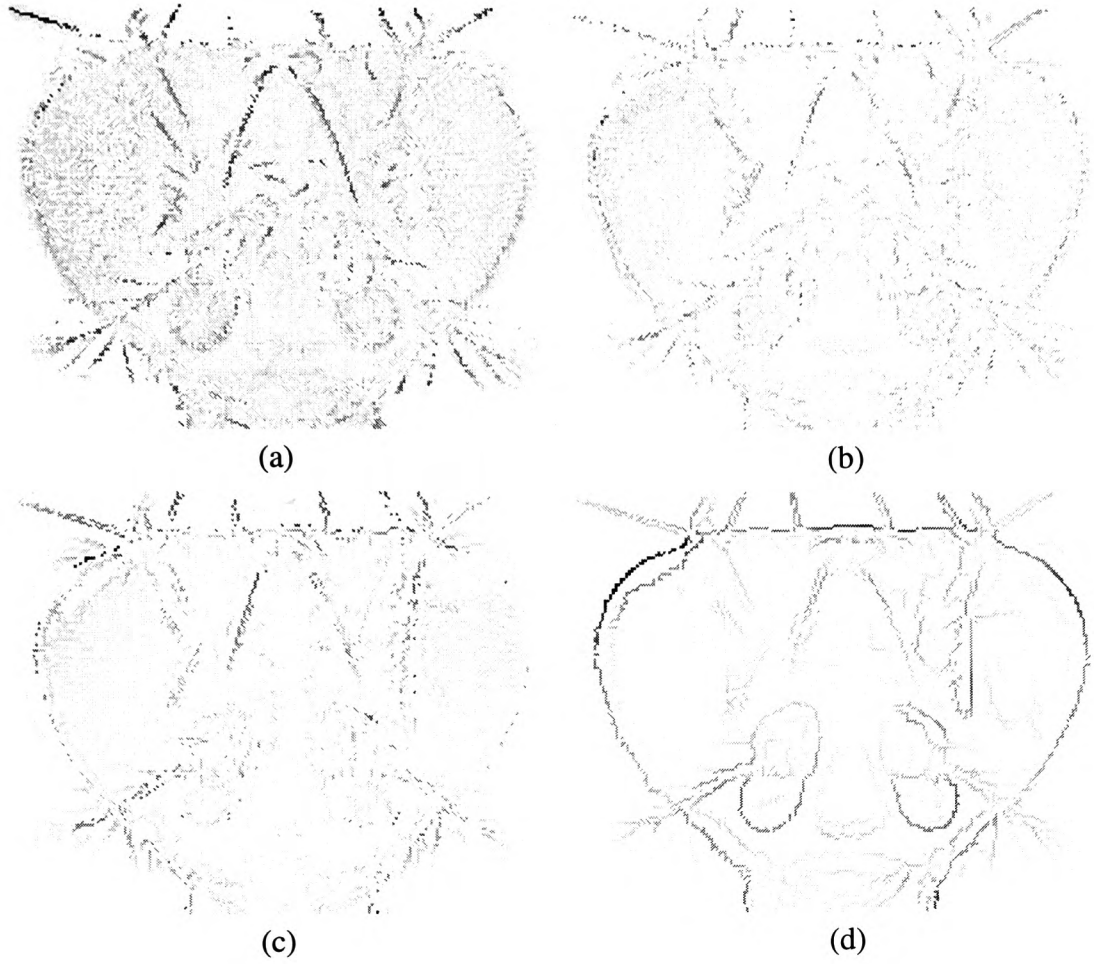


Figure 4



(a) (b)

(c) (d)

Figure 5

Figure 6

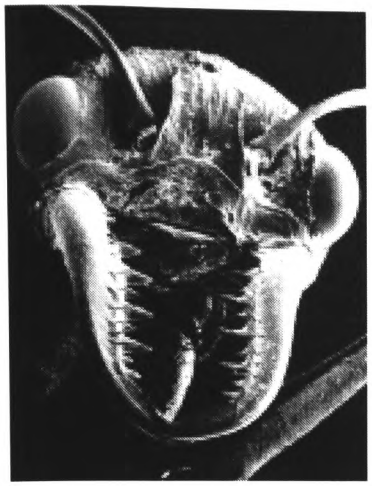
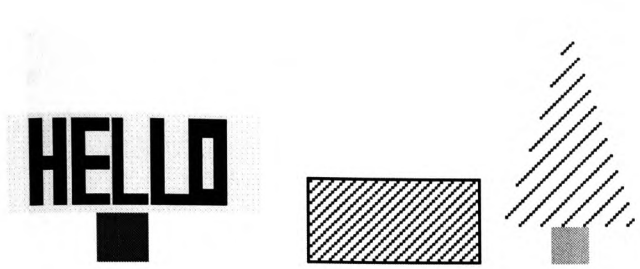
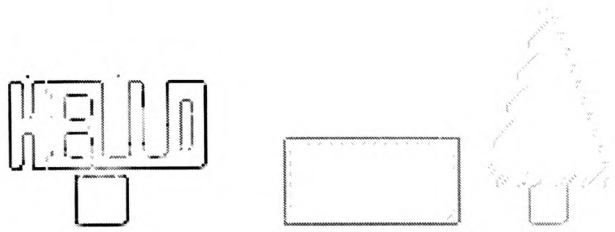


Figure 7

Figure 8

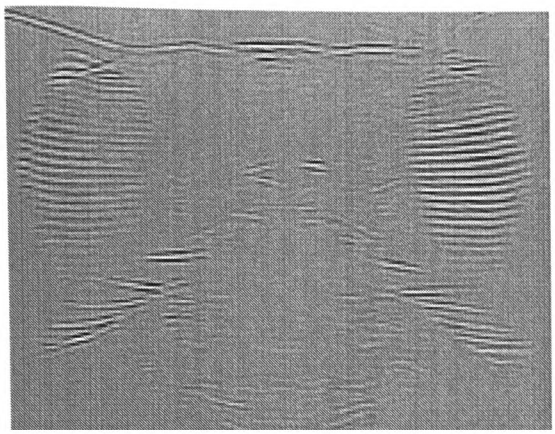


(a)

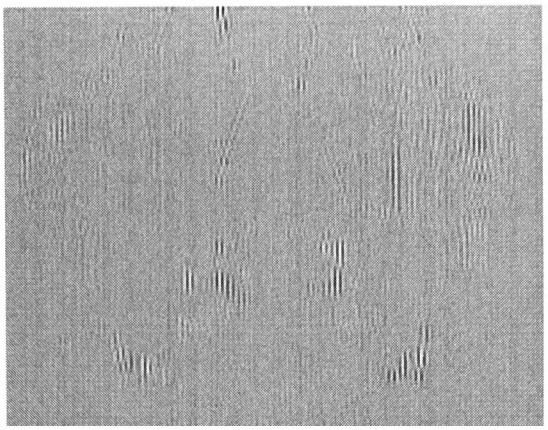


(b)

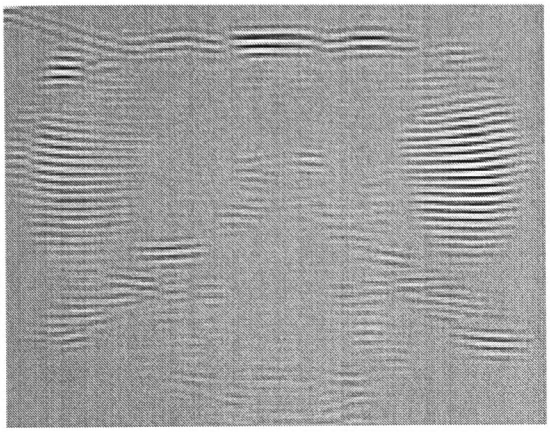
Figure 9



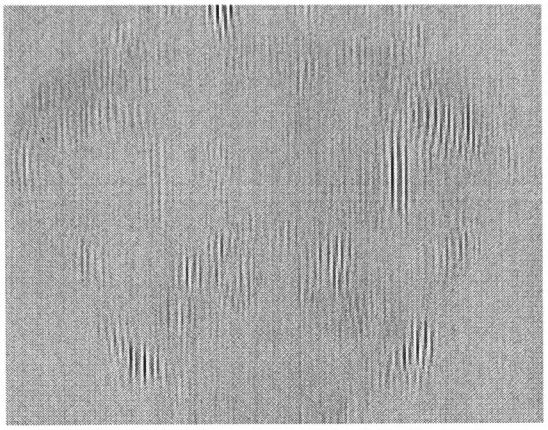
(a) Orientation 0 degrees;Scale S_2^2



(b) Orientation 90 degrees;Scale S_2^2



(c) Orientation 0 degrees;Scale S_2^3



(d) Orientation 90 degrees;Scale S_2^3

Figure 10

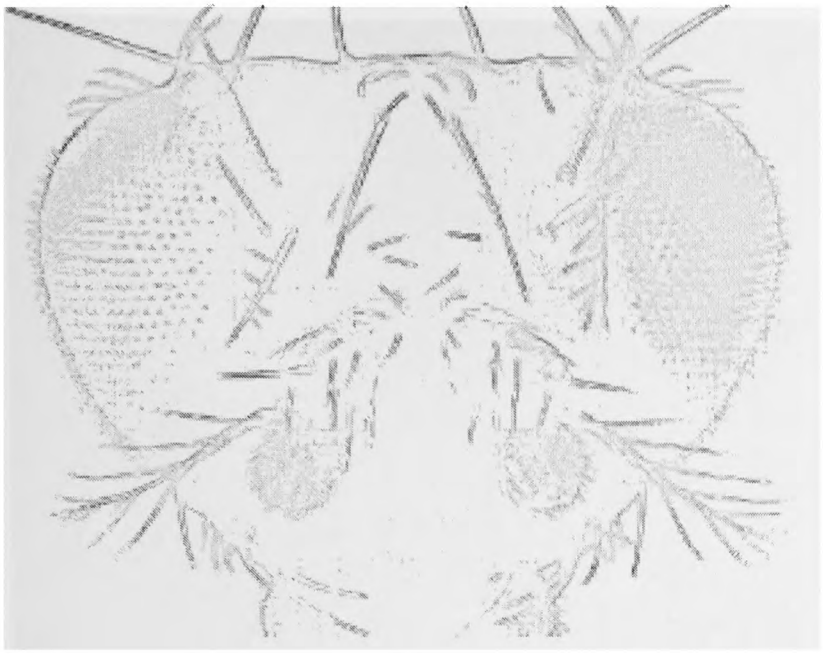


Figure 11

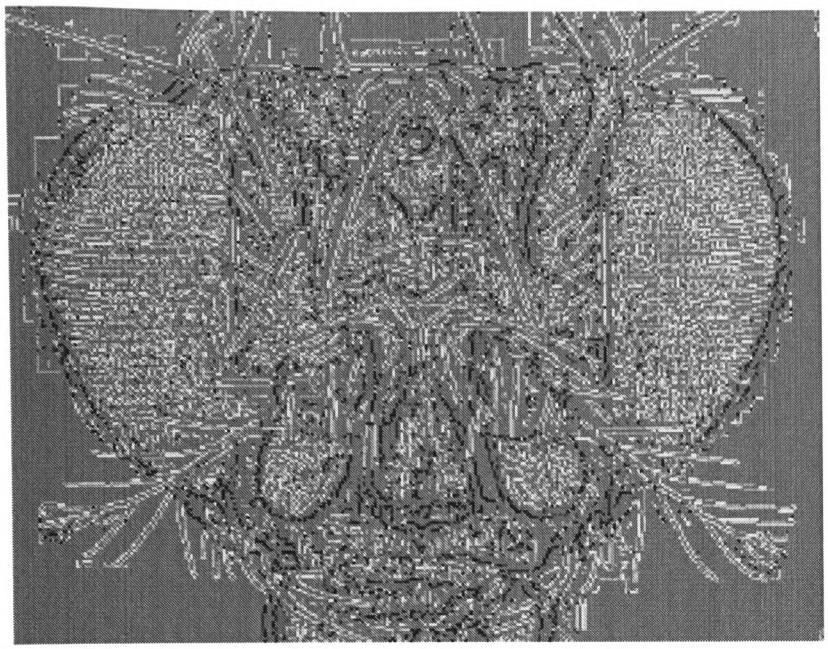


Figure 12